

From:  
Sent:  
To:  
Subject:

Chan, Christina  
Monday, February 11, 2002 4:06 PM  
Davis, Minh-Tam; STIC-Biotech/ChemLib  
RE: Rush search request for 09/775693

Please rush. Thanks Chris

-----Original Message-----

From: Davis, Minh-Tam  
Sent: Monday, February 11, 2002 3:34 PM  
To: Chan, Christina  
Subject: Rush search request for 09/775693

Please search in commercial database and in issued patent files:

1) SEQ ID NO:1 and a polypeptide encoded by SEQ ID NO:1. Please check to see any of said polypeptide is the same as SEQ ID NO:7 or 8.

Thank you.

MINH TAM DAVIS  
ART UNIT 1642, ROOM 8A01, MB 8E12  
305-2008

RECEIVED  
FEB 11 2002  
STIC

Point of Contact:  
Beverly Shears  
Technical Info. Specialist  
CM1 Tel: 305-4994  
1E05  
508-4994

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



## SEARCH REQUEST FORM

Requestor's  
Name: \_\_\_\_\_Serial  
Number: \_\_\_\_\_

Date: \_\_\_\_\_

Phone: \_\_\_\_\_

Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 02-13-02Searcher: Becker 4994Terminal time: 20

Elapsed time: \_\_\_\_\_

CPU time: \_\_\_\_\_

Total time: 25

Number of Searches: \_\_\_\_\_

Number of Databases: 1

## Search Site

STIC

CM-1

Pre-S

## Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

## Vendors

IG

STN

Dialog

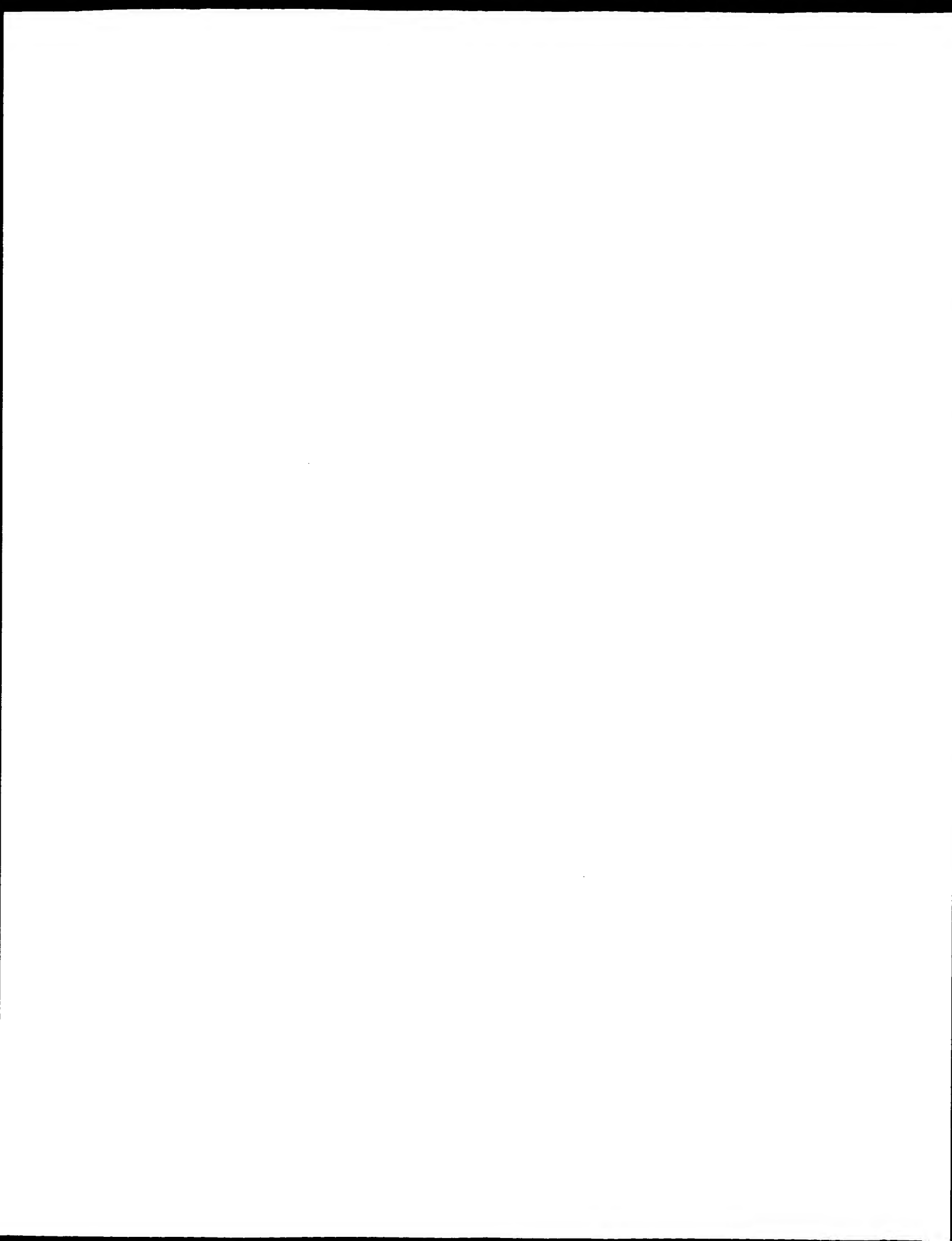
APS

Geninfo

SDC

DARC/Questel

☒ Other CGN





AC	
XX	13-MAR-2001 (first entry)
D7	
XX	Human prostate cancer antigen protein sequence SEQ ID NO:1594.
DE	
XX	Human prostate cancer; prostate cancer antigen; detection; diagnosis;
KM	neuroprotective; cyclostatic; cardioprotective; immunomodulatory; muscular;
KM	vulnerability; gastrointestinal; nephrotoxic; anti-infective; gynaecological;
KN	antibacterial; gene therapy; neural; immune; reproductive disorder;
KN	gastrointestinal; pulmonary; cardiovascular; gallbladder disease;
KN	wound; infectious disease.
XX	
OS	Homo sapiens.
XX	
PN	WO20005174-A1.
XX	
PD	21-SEP-2000.
XX	
PX	08-MAR-2000; 2000WO-US05988.
PF	
XX	12-MAR-1999; 99US-0124270.
PR	
XX	HUMAN GENOME SCI INC.

(HUMA-) ROSEN C A.  
PA  
PA (ROSE-)  
XX  
PI Rosen CA, Ruben SM;  
DR WPI: 2000-587513/55.  
XX N-PSDB: AAF16219.  
PT  
PT Prostate cancer associated gene sequences, referred to as prostate  
PS cancer antigens, useful for treatment, prevention, and diagnosis of  
PS disorders such as prostate cancer -  
PS  
PS Claim 11; Page 2041-2042; 238pp; English.

AAFI566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB563 to AAB57302,  
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC adriatic, antineoplastic, gynecological and antibacterial activities,  
CC and can be used as a gene therapy. The prostate cancer antigen, chromosome  
CC polynucleotides may be used for detection of numerous other diagnostic  
CC identification, as chromosome markers, and for numerous other therapeutic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, renal, and proliferative  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
CC  
CC  
XX Sequence 442 AA:  
SQ

alignment\_scores:  
Quality: 2123.00  
Ratio: 5.178  
Percent Similarity: 99.515

alignment\_block:  
US-09-775-693-1 x AAB57016 ..

Length: 412  
Gaps: 0  
Percent Identity: 99.272

Align seg 1/1

to: AAB57016 from: 1 to: 442

```

1 ATGTCGCAAGAGGCTCGTGTGTTGAGCTTACAGTGGGCGGCGAC 50
31 MetSerLeuYsgLysValValLeuAlaIysrLysLysLysLys 47
51 CTGCGCATCTCTGTGTGAGCAAGCAAGAGTATGAGCATTTGCT 100
47 rSerCysIleLeuValTrpLeuYsgLysLysLysLysLysLys 100
101 ATCGGCAACATTGGCCAGAGAGAGACTTCAGAGAGCCAGAGAG 150
64 yLeuAlaAsnIleGlyLysLysLysLysLysLysLysLysLys 150
151 GCAGTGAAGCTTGGGCAAGAGAGTTCATGAGATGTCAGAGAG 200
81 AlaLeuYsgLysLysLysLysLysLysLysLysLysLysLys 200
201 GTTGTGAGAGTTCATGAGTGGCGGCGCATCTGAGCGGCGAGT 250
97 upheValGluGluIupheIleTrpProAlaIleIleIleIle 250
251 AGGACCGCTACCTCTGCGACCTCTTGGCCAGAGCGGCGATG 300
114 IuAspArgTrpLeuLeuGlyTrpSerLeuAlaArgProCysIle 300
301 AAACAGTGAAGTTCGCGGCGAGGCGGCGCAAGTATGTCGCG 350
131 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTrpVal 350
351 CGGCAGAGAGAGAGAGAGAGTCCGCTTGGAGTTCAGCTGCT 400
147 yAlaThrIleLysGlnLysAspGlnAlaTrpPheGlnLeuSer 400
401 CACTGGGCGGCGAGTAAAGTTCATGCTCTGAGAGAGCGTAA 450
164 erLeuAlaProGlnIleLysValIleAlaProTrpArgPhePro 450
451 TACAGCGGTTCAAGGCGGCAATGAGTGAAGTTCGAGAGAG 500
181 TyrAsnArgPheLysGlnAlaGlnAspLeuMetGlnTrpAla 500
501 CGGATTCCTACCTCGGCTCCAGAGAGCGTGGAGAGAGAGAG 550
197 sGlyIleProIleProValTrpProLysAsnProTrpSerMet 550
551 ACCTCATGCAATGAGTTCAGAGAGTGGAGTTCGAGAGAGAG 600
214 sLeuMetHisIleSerTrpGlnAlaGlnIleLeuGlnAsnPro 600
601 CAAGCGGCTCCAGGTTCTTACAGAGAGAGAGAGAGAGAGAG 650
231 GlnAlaProProGlyLeuTrpThrLysThrLysProAlaLys 650
651 CAACAGCGGCTTCAGATGAGTTCGAGTTCGAGTTCGAGTTC 700
247 oAsnThrProAspIleLeuGlnIleGlnPheLysLysGlnVal 700
701 AGGTGACCACTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAG 750
264 yValTrpAsnValLysAspGlyTrpThrHisGlnTrpSerLeu 750
751 TTTCATGTTACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800
281 PheMetTrpLeuAsnGlnValAlaGlnLysLysLysLysLys 800
801 CATCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 850
297 rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyLys 850
851 CCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

```

us-09-775-693-1.fag

```

314 hrrProAlaGlyTrpIleLeuTrpHisAlaHisLeuAspIleGlnAlaPhe 330
901 ACCATGACCGGCAAGTGGCGGCAAAATCAAAACAGAGCTGGCTGAAT 350
331 ThrMetAspArgGlnValAlaTrpLysLysLysLysLysLysLys 350
951 TGCTGAGCTGTGTATACCGGTTTACCGGCTACCGCTGAGTGAATTTG 400
347 eAlaGlnLeuValLysTrpGlyPheTrpHisSerProGlnLysLys 400
1001 TCGGCGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
364 aIArgHisCysIleAlaLysSerGlnGlnAlaGlnLysLysLys 450
1051 GTGTCGCTCTCAAGGCGGAGAGTTCATGCTGCGGCGGAGTCCGCACT 500
381 ValSerValLeuLysGlnValTrpIleLeuGlyAlaGlnSerPro 500
1101 GTCTCTTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
397 uSerLeuTrpAsnGlnGlnLysValSerMetLysValGlnLysLys 550
1151 ACCCACTGATGCCAGCGGCTTCATCAACATCAATTCCTGAGCTGAAG 600
414 IupProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuAla 600
1201 GAATATCATCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 650
431 GluTrpHisArgLeuGlnSerLysValThrAlaLys 650
seq.name: /SIPS2/gcgdata/geneseq/genesep/AA2001.DAT: AAG75416
seq_documentation_block:
ID AAG75416 standard; Protein: 442 AA.
AC AAG75416;
DN 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein SEQ ID NO:6180.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection.
XX colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX PF 28-SEP-2000; 2000MO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX DR WPI: 2001-235357/24.
XX N-PSDB: AAH34821.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7633-7624; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens (P), where
XX cancer antigens have cytostatic activity and can be used in gene
XX diagnosis and treatment of diseases associated with inappropriate p

```

AA	Sequence	442 AA;
50		

alignment\_block: 1

1 ATGTCAGCAAGCGCTCCGTGGTTCACGCCAACATCAGG

```

214 snlwmethistilesertygluialy1leuenglaspnProlyAsn 230
601 CAAGGGCTTCAGGCTCTACACGAGACCCAGACCCAGCCAAAGCCCC 650
|||||
231 glialaprobioyleutyThryThryThrglnaspproialySalapI 247
|||||
651 CACACCCCTGACATTCCTGAGATCGAGTTCAAAAAGGGTCCCTGTGA 700
|||||
247 oasnthpPrcasprileuenglillegluphelnlysglYvalProvalL 264
|||||
701 AGGTGACACACGTCAAGATGTGACACACCCACACAGACTCCCTTGGAGCTC 750
|||||
264 yvalIthnasvalysaspglYlhrthrhnsGlnthrsleugluleu 280
|||||
751 TTCATGTACSTGAACGAACTCCGGGCGCAAGCATGAGCTGGCGGTATTGA 800
|||||
281 Rhemetyrleuasnlglnvalalaglyushnsglyalaglyatgltleas 297
|||||
801 CATGTGGGAGAACCCGTTCATTGGAAATGAAGTCCCGAGATTCCTACGAGA 850
|||||
297 rtlevalglusnarpheileglymetussseirpglYtlelyglut 314
|||||
851 CCCGACGACACATCCCTTACCAAGTCATATTGACATCGAGACCTTC 900
|||||
314 hrProialglYhrThleuuythlsAnhlsleuaspIleclYalAphE 330
|||||
901 ACCATGAGACCGGAGAGTCCGCAAAATCAACAAGCCCTGGCTGTAAATT 950
|||||
331 ThmetasprtggluValatrglys1leu5glnglyleu6glyleu15ph 347
|||||
951 TGTGTAGCTGGTGTATACCGGTTTACGGCCCTAGCCCTGAGTGTGAATTG 1000
|||||
347 ealnglYleuValtyThrglyhrpethrhlssepproglyucysgluphev 364
|||||
1001 TCGGCACTGCATCGCCAAAGTCCACAGACGAGTGGAGGAAGGAAAGTCAg 1050
|||||
364 alayrhnsCysllealalyserglnlglnatrgvalglnglylYvalgln 380
|||||
1051 GTGTCCGCTCTCAAGGGCCAGGCTGTCATCTCCGCGGAGTCCCACT 1100
|||||
381 valserValleuylsGlygluValYlTrlIleuenglYatrglYserProle 397
|||||
1101 GTCTCTACATATGAGAGAGCTGTGAGCATGAACCTGCAAGGTCATTATG 1150
|||||
397 useIleuYlrgasnlglnleuValsermetasnvalglnglYsprlycg 414
|||||
1151 AGCCAACTGATGCCACCGGGGTTCATCAACATCAATTCCCTCAGGCTGAAG 1200
|||||
414 lnrProthrasplathrlylphelIleasnIleasnserleuayrleuyls 430
|||||
1201 GAATATCATCTCTCCAGAGCAAGGTCACGTCCAAA 1235
|||||
431 gluyrthlsayrleuGlInserlysalIthralYals 442
|||||
seq_name: /SID52/gcadata/geneseq/geneseq/AA2001.DAT:AAg82054
seq_documentation_block:
ID AAg82054 standard; Protein: 367 AA.
XX
AC AAg82054:
XX
DT 03-SEP-2001 (first entry)
XX
DE 5. epidermidis open reading frame protein sequence SEQ ID NO:1202.
XX
XX Staphylococcus epidermidis SRI strain; Infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
ID 17-MAY-2001.

```

XX 09-NOV-2000; 2000WO-US30782.  
 XX 09-NOV-1999; 99US-0164258.  
 XX (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX  
 XX  
 PI  
 XX  
 DR WPI: 2001-316495/33.  
 DR N-PSDB: AAH52904.  
 XX  
 PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 useful for vaccinating against infections, e.g. endocarditis.  
 XX  
 PS Claim 18; Page 345; 2188BP; English.  
 XX  
 CC AAH5304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*.  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC in the exemplification of the present invention. AAH55091 to  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC no sequences are present for the disclosure for SEQ ID NO:4465 to 4472.  
 XX  
 50 Sequence 367 AA:

alignment\_scores:  
 Quality: 747.00 Length: 373  
 Ratio: 2.756 Gaps: 3  
 Percent Similarity: 72.654 Percent Identity: 40.214

alignment\_block:  
 US-09-775-693-1 x AAH82054 ..

Align seg 1/1 to: AAH82054 from: 1 to: 367

```

10 AAAGGCTCCGTTGTTCTGCTTACAGTGGCGGCTGACACCTGTCAT 59
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
  2 LysAspLysIleValIleuAlaTyrSerGlyLeuAspThrValAla 18
   :|: |||||: :|||:|||||:|||||:|||||:|||||:|||||:
  60 CCTGCTGCTGCTGAAGGAACAAGCTATGACGTTCATTCCTATGCGCA 109
   :|: |||||: :|||:|||||:|||||:|||||:|||||:|||||:
  18 aValGIdTrpIleuIleAspLysGlyTyrAspValValAlaIleAsp 35
   :|: |||||: :|||:|||||:|||||:|||||:|||||:|||||:
  110 ACATTGGCCCAAGGAAGACTTGCAGAGCCGAGGAAGAGCCTGAAG 159
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  35 sPvalGlyGluGlyLysAspLeuAspValValTyrGlnLysAlaLeuAsp 51
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  160 CTGGGGCCCAAAAGGTTCATGAGAGTGTGACAGGAGGATTGTGTA 209
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  52 MetGlyAlaValGluGlyHisIleIleAspAlaThrLysGlnPheSerAs 68
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  210 GGAGTTCATCTGGCCGCTACCTGACGCGACCTGATGATGAGACGCT 259
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  68 pAspTyrValIleTyrAlaIleLysGlyAsnLeuMetTyrGlnAsnAla 85
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
  260 ACCCTCTGGGACACTCTCTGCAAGCCCTGTCATCCCGCAAAAGAGTG 309
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:

```

```

85 YrProLeuValSerAlaLeuSerArgProLeuIleAlaLysLysLeuVal 101
310 GAATTCGCCCGAGCGGAGGAGGCGCAAGTATGTGTCGCCACGCGCCACAG 359
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
102 GluIleAlaGluLysThrAsnSerIleGlyIleAlaHisGlySerThrG 118
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
360 AAAGGGAACGATCAGATCGCGGTTCAGCTCAGCTGCTACTGACCTGGCC 409
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
118 YLysGlyAsnAspGlnValArgPheGluValAlaIleLysAlaLeuAsp 135
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
410 CCGAGTAAAGGTTCATTCCTCCCTGAGAGATCCCTGAATTCACAACCG 459
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
135 rOlyLysLeuLysAlaPheAlaProValArgGluTrpAlaTrp..... 148
460 TTCAGGCGCGCAATGACCTGATGCTACGCAAGACACGCGATTC 509
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
149 .....SerArgGluGluGluIleAspTyrAlaIleLysHisAsnIlePr 163
510 CATCCCGCTACCTCCCAAGACCGGTGACCATGATGAGACCTGATGC 559
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
163 oValSerIleAsnTyrAspSerProTyrSerIleAspGlnAsnLeuTyrG 180
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
560 ACATCAGCTACGAGCGTGAATCTGAGAACCCCAAGACCAACGCGCT 609
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
180 LysArgAlaAsnGluGlyGlyIleLeuGluAspProTyrAlaAlaProPro 196
610 CCAAGTCTTACACGAAGACCCAGACCCAGCAAGAGCCCAACACCC 659
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
197 GluAspAlaPheAspLeuThrProLeuGluGluTrpProAspAla 213
660 TGACATTCCTCAGATCGAATTCATAAAGGCGCTGTGAGTGACCA 709
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
213 AspArgIleIleLeuThrPheLysGlnGlyIleProValGlnVal..... 228
710 ACGTCAGAGTGGCACCCACCAAGACCTCTGGAAGCTTCATGATAC 759
   |||||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
229 .....AspIlyLysAspTyrGlnLeuAsp...AspLeuIleLeuTyr 241
760 CTGAAGAGTCCGCGGCAAGCATGCGCGGCGGCGATTCGACATCGTGA 809
   |||||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
242 LeuAsnGlnIleuAlaGlyLysHisGlyIleGlyArgIleAspHisValAl 258
810 GAACCGCTCATGATGATGAGTCCGAGGTATCTACGAGACCCAGAG 859
   |||||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
258 uAsnArgMetValIleLysSerArgGluIleTyrGlnThrProGlyA 275
860 GCACATCTTTACATGCTCATTTAGACATGAGAGCCCTTACCATGAC 909
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
275 IacIuValIleLeuLysAlaHisLysAlaLeuGluThrIleThrLeuTyr 291
910 CGGAAAGTGGCGCAAAATCAACAAGCCCTGGAATTGCTGAGCT 959
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
292 LysAspValAlaHisPheLysProValIleGluLysGlnPheSerGlu 308
960 GGTGTATACCGGTTTACGCGCTACCGCTGAGTGAATTTGTCGCCACT 1009
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
308 nIleTyrAsnGlyLeuTrpPheSerProLeuThrAspSerLeuLysLeuP 325
1010 GCATGCGCAAGTCCCGAGGAGTGCAGAGGAAGTGCAGGTGTCGTC 1059
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
325 heIleAspSerThrGlnGlnTyrValGlnIleLysPheArgIleLysLeu 341
1060 CTCAAGGCGCCAGGTGTATCATCTCGCGCGGAGTCCCGCTCTCTCTTA 1109
   :|||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
342 PheLysGlyAsnAlaIleValaIleGlnGlyAsnGlnTyrThrLeuTyr 358
1110 CAATGAGAGCTGTGAGGC 1128
   |||:|||||: :|||:|||||:|||||:|||||:|||||:|||||:
358 rAspGluLysLeuAlaThr 364

```

seq\_name: /std2/gcgdata/geneseq/geneseq/AA2001.DAT:AA91292  
 seq\_documentation\_block:



```

322 erleuaspalaheileaspserthrcingluhishalvthrclyasp1le 338
1048 CAGGTGTCCTGCTCAAGGCCGAGGTTCATCTCCGCCGAGCTCC 1097
339 ArgMetValLeuHisIaIcylserIleThrIleasnGlyArgArgSerSe 355
1098 ACTGTCCTCTACATAGAGAGCTGTGACATGACGTGACAGGTCATT 1147
355 rHisSerLeuTyrAspPheasnLeuAlaIleThrTyrAspThrTyr 372
1148 ATGAGCCACTGATGACCGGTCATCAACATCAATCCCTCAGGCTG 1197
372 heaspIleThrLeuAlaIlysglyPheValGlnLeu..... 383
1198 AAGGATATTCATCTCCAGCAGCAAGCTCAGCAAA 1236
384 .....HisGlyLeuSerSerLysIleAlaasnLys 393

```

seq\_name: /SID2/gcgdata/geneseq/AA2001.DAT: AAB79801  
seq\_documentation\_block:

ID AAB79801 standard; Protein: 401 AA.

AC AAB79801;

30-APR-2001 (first entry)

Corynebacterium glutamicum MP protein sequence SPQ ID NO:336.

Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
fine chemical production; microorganism; organic acid; nucleoside;  
nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
carbohydrate; aromatic compound; cofactor; polyketide; enzyme.

Corynebacterium glutamicum.

MO200100843-A2.

04-JAN-2001.

23-JUN-2000; 2000WO-IB00923.

```

25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.

```

```

PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 31-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.

PA (BAD1 ) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX
XX MPI: 2001-137957/14.
XX
XX N-PDB: AAF71920.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX pathway proteins, useful for producing fine chemicals in
XX PT microorganisms, including organic acids, nonproteinogenic amino
XX PT and purine and pyrimidine bases -
XX
XX Claim 20; Page 656-657; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX MP nucleic acids are useful for the production of fine chemicals
XX CC in microorganisms, including organic acids, nonproteinogenic amino
XX CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 401 AA;

```

#### alignment\_scores:

Quality:	619.50	Length:	413
Ratio:	2.261 <td>Gaps:</td> <td>8</td>	Gaps:	8
Percent Similarity:	66.344	Percent Identity:	35.351

#### alignment\_block:

US-09-775-693-1 x AAB79801 ..

Align seg 1/1. to: AAB79801 from: 1 to: 401

```

19 GTGGTCTGSCCTACAGATGGCGGCTGGACACCTGCTCATCTCGGTG 68
:::|||||
5 IleValLeuAlaTyrSerGlyLeuAspThrValAlaIleProGly 21
:::|||||
69 GCTGAAGAGAACAA...GGCTATGACGTGATGCGCTATGCGCAACATG 115
:::|||||
21 rLeuLysLysMetIleAspGlyGluValIleAlaValSerLeuAspLeuG 38
:::|||||

```

[illegible]

```

::: ||| .....|||.....||| ::::
322 etLeuAspAlaPheIleaspSerThrGlnHisValThrGlyAspIle 338
1048 CAGGGTCGCCGTCCACGAGCGCAAGCTCATCCTCGCCGCGAGATCCC 1097
:::::::::: |::|::|::|::|::|::|::|::|::|::|::|::|::|
339 ArgMetValLeuHisAlaGlySerIleThrIleasnGlyAcyArgSerSe 355
1098 ACTGCTCTCTACAATTGAGAGAGTGCAGCATGAACGTCGAGGTCGATT 1147
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
355 rHISserLeuTyraSpHeasneLeuAlaInTrTyraSprthrGlyAspThr 372
1148 ATAGACCACCTGATGTCACCGGGTTTCATCAACTCAATTCCCTCAGGCTG 1197
372 heAsprGlInThrLeuAlalysglYpheValGlnLeu..... 383
1198 AAGCAATATCATGCTCTCCAGCACGATGCTACGCCCAA 1236
384 .....HisIcylseuSerIstyIleLaAnalys 393

seq_name: /SLDS2/gcgdata/geneseq/geneseqp/Aa2000.DAT:AAG08821
seq_documentation_block:
ID AAG08821 standard; Protein: 302 AA.
XX
AC AAG08821:
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6508.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD
XX
PF 06-SEP-2000.
PR
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130444.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
```



PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 14-JUN-1999; 9905-0138847.  
PR 16-JUN-1999; 9905-0139119.  
PR 17-JUN-1999; 9905-0139452.  
PR 18-JUN-1999; 9905-0139453.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 21-JUN-1999; 9905-0139763.  
PR 22-JUN-1999; 9905-0139817.  
PR 23-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140351.  
PR 24-JUN-1999; 9905-0140354.  
PR 28-JUN-1999; 9905-0140695.  
PR 29-JUN-1999; 9905-0140823.  
PR 30-JUN-1999; 9905-0140991.  
PR 01-JUL-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 02-JUL-1999; 9905-0142154.  
PR 06-JUL-1999; 9905-0142055.  
PR 08-JUL-1999; 9905-0142380.  
PR 09-JUL-1999; 9905-0142803.  
PR 12-JUL-1999; 9905-0142920.  
PR 13-JUL-1999; 9905-0143427.  
PR 14-JUL-1999; 9905-0143542.  
PR 15-JUL-1999; 9905-0143624.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 21-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 23-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 26-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 9905-0146388.  
  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 06-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 10-AUG-1999; 9905-0147935.  
PR 11-AUG-1999; 9905-0148171.  
PR 12-AUG-1999; 9905-0148319.  
PR 13-AUG-1999; 9905-0148565.  
PR 16-AUG-1999; 9905-0148684.  
PR 17-AUG-1999; 9905-0149368.  
PR 18-AUG-1999; 9905-0149175.  
PR 20-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 23-AUG-1999; 9905-0149929.  
PR 25-AUG-1999; 9905-0149930.  
PR 26-AUG-1999; 9905-0150566.  
PR 27-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 30-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 01-SEP-1999; 9905-0151438.  
PR 07-SEP-1999; 9905-0151930.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154038.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 28-SEP-1999; 9905-0155659.  
PR 29-SEP-1999; 9905-0156458.  
PR 05-OCT-1999; 9905-0156596.  
PR 06-OCT-1999; 9905-0157753.  
PR 07-OCT-1999; 9905-0157865.  
PR 08-OCT-1999; 9905-0158029.  
PR 12-OCT-1999; 9905-0158232.  
PR 13-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 18-OCT-1999; 9905-0159638.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160814.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.



28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality: 533.00 Length: 203  
Ratio: 3.270 Gaps: 3  
Percent Similarity: 80.296 Percent Identity: 50.246

## alignment\_block:

US-09-775-693-1 x AAG08821 ..

Align seg 1/1 to: AAG08821 from: 1 to: 302

```

19 GTGCTTCTGGCCATGAGTGGGCGGCGTGCAGACCTGCTGCTGCTG 68
   |||||
99 ValValLeuAlaIleValSerGlyLeuAspThrSerValIleValProtr 115
69 GCTGAAGAACAA...GGCTATGACGTATGCTGCTGCTGCTGCTG 115
   |||||
115 PleuLysGluAsnTyrGlyCysGluValValCysPheThrAlaAspValG 132
116 GCCGGAAG...GAAGACTTGAGAGAACCCGAGAACGACCTGAGCTT 162
   |||||
132 LysGlnGlyLeuLysGluLeuGlnLeuGlnGlnLysAlaLysAlaSer 148
163 GGGGCAAAAGGTTCATTGAGATGTCAGCAGGAGGATTGTCAGCA 212
   |||||
149 GlyAlaSerGlnLeuValAlaAsnAspLeuThrGlnLeuValLysAs 165
213 GTTCATCTGGCCGCGCATCCAGTCCAGCGACCTGATGAGCGGCTAC 262
   |||||
165 PheIlePheProCysLeuArgAlaGlyAlaIleTyrGlnArgLysTyr 182
263 TCCCTGGGACCTCTCTGTCAGGCGCTGTCATGCGCCGCAACAGTGA 312
   |||||
182 euLeuGlyThrSerMetAlaArgProValIleAlaLysAlaMetValAsp 198
313 ATCGCCCAAGGGAGGGGCGCCAGATATGTCGCCAGCGGCGCACAGAA 362
   |||||
199 ValAlaIleGlnValGlyAlaAspAlaValAlaIleGlyCysThrGly 215
363 GGGGACGATGACGATGCGGCTTGGACCTGACCTGCTGCTGCTGCTG 412
   |||||
215 SerGlyAsnAspGlnValAlaArgPheGlnLeuThrPheSerLeuAsnProG 232
413 AGATAAAGGTGATGCTCCCTGAGAGATCCCTGATTCACAAACGGGTT 462
   |||||
232 IuLeuLysValIleAlaProtrPargGlnLysPheGln.....Ile 244
463 AAGGCGCCGAATGACCTGATGAGTGCAGCAACGACACAGCGGATTC 512
   |||||
245 GlnGlyArgGlnAspAlaIleGlnTyrAlaLysLysHisAsnValProva 261
513 CCCGTCACCTCCCAAGAACCCGTGAGCATGATGAGAACCTCATGCA 562
   |||||
261 IProValThrLysSerIleTyrSerIleTyrSerIleTyrSerIleTyr 278
563 TCAGCTACGAGGCTGGAATCTGAGAACCCCAAGAACGCGCTTCA 612
   |||||
278 euSerHisGlnGlyAspLeuLysGlnAspProAlaAsnGlnLysPhe 294
613 GGTCTCTTAC 621
   |||||
295 Aspmetyr 297
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB53535
seq_documentation_block:
ID AAB53535 standard; Protein: 146 AA.
XX
AC
XX AAB53535:
XX

```

09-MAR-2001 (first entry)

Human colon cancer antigen protein sequence SEQ ID NO:1075.

Human: colon cancer; colon cancer antigen; diagnosis; detection;  
identification; cytostatic; cardioactive; neuroprotective; vulnary;  
immunomodulatory; muscular; gynecological; gastrointestinal;  
neurotropic; antineoplastic; antibacterial; gene therapy; wound;  
neural disorder; immune system disorder; muscular disorder;  
reproductive disorder; gastrointestinal disorder; renal disorder;  
infectious disease; cardiovascular disorder.

Homo sapiens.

WO20005351-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-050583.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-587534/55.

N-PSDB; AAC98292.

COLON cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -

Claim 11: Page 1660-1661; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
called human colon cancer antigens, given in AAB53334 to AAB54006. The  
human colon cancer antigens can have cytostatic, cardioactive, muscular;  
neuroprotective, immunomodulatory, gynecological, gastrointestinal, and  
vulnary; nephrotropic, antineoplastic, antibacterial activities, and  
can be used in gene therapy. The colon cancer antigen polynucleotides,  
proteins and antibodies of the proteins are useful for the prevention,  
treatment and diagnosis of colon disorders, such as colon cancer. The  
polynucleotides may be used in diagnostics and research, such as for  
chromosome identification, and as hybridization probes. The proteins  
may also be used to prevent diseases such as neural disorders, immune  
system disorders, muscular disorders, reproductive disorders,  
gastrointestinal disorders, wounds, renal disorders, infectious  
diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
invention.

Sequence 146 AA:

alignment\_scores: Quality: 463.00 Length: 99

Ratio: 4.926 Gaps: 0

Percent Similarity: 94.949 Percent Identity: 92.929

alignment\_block:

US-09-775-693-1 x AAB53535 ..

Align seg 1/1 to: AAB53535 from: 1 to: 146

```

838 GGTATCTACGAGACCCAGCAGCAGCATCTTACCATGCTGCTTAGA 887
   |||||
1 GlyThrSerGlnTyrProAlaGlyThrIleLeuTyrHisAlaHisLeuAs 17
888 CATGAGGCTTCACCATGAGACCGGAGAGTCCGCAATCAACACAGGCC 937
   |||||
17 PileGlnAlaPheThrMetAspArgGlnValAlaArgLysIleLysGlnGln 34

```

```

938 TGGGCTGAATTTGCTGAGCTGTGTATACGCGTTAGCGCTACCT 987
|||||
34 enGlyLeuLysPheAlaGlu***ValTyrThrGlyPheTrpHisSerPro 50
|||||
988 GAGCGTGAATTTGCTGCGCCAGTCGATGCCAAGTCCCAAGACGAGTGA 1037
|||||
51 GlucylSerIlePheValArgHisCysIleAlaLysSerGlnGluArgValGln 67
|||||
1038 AGGGAAGTGCAGGTGCTCCGCTCAGCGCCAGGTGATACCTCCGCGC 1087
|||||
67 uGlyLysValGlnValSerValLeuLysGlyGlnValTyrIleLeuGlyA 84
|||||
1088 GGGAGTCCCGACTGCTCTCTCAATGAGGAGCTGCGAGCATGAA 1134
|||||
84 rGlySerProLeuSerLeuTyrAsnGlnGluLeuValSerMetLasp 99
|||||
seq_name: /SIDS2/gcycdata/geneseq/geneseq/AA2001.DAT:AA16101
seq_documentation_block:
ID AA16101 standard; Protein; 59 AA.
AC AA16101:
XX
XX
XX 12-OCT-2001 (first entry)
DE Peptide #2535 encoded by probe for measuring cervical gene expression.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX Homo sapiens.
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR.
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID No 20927; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SNP: see A110068-A128459). The present sequence is a peptide encoded
XX by one such probe. The SNPs are derived from human HeLa cells. The SNPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 59 AA:

```

alignment\_scores:

```

Quality: 228.00 Length: 55
Ratio: 4.560 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 76.364
alignment_block:
US-09-775-693-1/rev x AA16101
..
Align seg 1/1 to: AA16101 from: 1 to: 59
573 CTCGAGCTGATGTCATGAGTTCTCAWCCATGCTCCAGCGGTTCTGG 524
|||||
1 LeuIleAlaAspThrHisGluValLeuValHisAspProArgValLeuG1 17
523 GAGTACCGCGGATGGGATCCCGTGTGCTTGGCTACCTCCATCAGGTCA 474
|||||
17 userAspIrrAspGlyAsnProMetLeuLeuCyValPheHisGlnIle1 34
473 TTGGCGCCCTTGAAACCGTGTGATGATTCAGGATCCCTCCAGGAGCAAT 424
|||||
34 LeuIleAlaIleGluProValValGluLeuGlnLysHisProGlySerAsn 50
423 GACCTTATCTGGGG 409
|||||
51 AspleuAsnLeuGly 55
|||||
seq_name: /SIDS2/gcycdata/geneseq/geneseq/AA2001.DAT:AA28592
seq_documentation_block:
ID AA28592 standard; Protein; 59 AA.
AC AA28592:
XX
XX
XX 17-OCT-2001 (first entry)
DE Peptide #2629 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX Homo sapiens.
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR.
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX Claim 27; SEQ ID No 28861; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SNP:
XX see A113115-A1157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.

```

SQ Sequence 59 AA:

alignment\_scores:

Quality: 228.00 Length: 55

Ratio: 4.560 Gaps: 0

Percent Similarity: 90.909 Percent Identity: 76.364

alignment\_block:

US-09-775-693-1/rev x AAM28592 ..

Align seg 1/1 to: AAM28592 from: 1 to: 59

573 CTCGTAGCTGATGTCATGAGGTTCTCATGCTCCACGGGTTCTTG 524  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1 LeuIleAlaAspTrpHisGluValLeuValHisAspProArgValLeuGI 17  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 523 GAGTGACCGGATGGGAATCCCGTGTGCTTGGCTACCTCCAGGAGCAAT 474  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 17 uSerAspTrpAspGlyAsnPrometLeuLeuGlyValPheHisGlnIleI 34  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 473 TTGCGGCCCTTGACCGGTTGTAGATTCAGGCAATCCTCCAGGAGCAAT 424  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 34 LeuIleAlaIleGluProValAlaGluLeuGlyHisProProGlySerAsn 50  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 423 GACCTTTATCTGGGG 409  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 51 AspLeuAsnLeuGly 55

seq\_name: /SIDS2/gcgcdata/geneseq/geneseqp/AA2001.DAT: AAM03828

seq\_documentation\_block:

ID AAM03828 standard; Protein: 59 AA.

AC AAM03828:

XX 09-OCT-2001 (first entry)

DE Peptide #2510 encoded by probe for measuring breast gene expression.

KW Probe: human; breast disease; breast cancer; development disorder.

XX Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001: 2001MO-US00661.

XX 04-FEB-2000: 2000US-0180312.

XX 26-MAY-2000: 2000US-0207456.

XX 30-JUN-2000: 2000US-0608408.

XX 03-AUG-2000: 2000US-0632366.

XX 21-SEP-2000: 2000US-0234687.

XX 27-SEP-2000: 2000US-0236359.

XX 04-OCT-2000: 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR.

XX WPI: 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX Claim 27: SEQ ID No 12568; 322pp: English.

XX The present invention relates to novel single exon nucleic acid probes

XX (see A100010-A110067). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for measuring human gene expression in

CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 59 AA:

alignment\_scores:

Quality: 228.00 Length: 55

Ratio: 4.560 Gaps: 0

Percent Similarity: 90.909 Percent Identity: 76.364

alignment\_block:

US-09-775-693-1/rev x AAM03828 ..

Align seg 1/1 to: AAM03828 from: 1 to: 59

573 CTCGTAGCTGATGTCATGAGGTTCTCATGCTCCACGGGTTCTTG 524  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 1 LeuIleAlaAspTrpHisGluValLeuValHisAspProArgValLeuGI 17  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 523 GAGTGACCGGATGGGAATCCCGTGTGCTTGGCTACCTCCAGGAGCAAT 474  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 17 uSerAspTrpAspGlyAsnPrometLeuLeuGlyValPheHisGlnIleI 34  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 473 TTGCGGCCCTTGACCGGTTGTAGATTCAGGCAATCCTCCAGGAGCAAT 424  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 34 LeuIleAlaIleGluProValAlaGluLeuGlyHisProProGlySerAsn 50  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 423 GACCTTTATCTGGGG 409  
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 51 AspLeuAsnLeuGly 55

seq\_name: /SIDS2/gcgcdata/geneseq/geneseqp/AA2001.DAT: AAB59826

seq\_documentation\_block:

ID AAB59826 standard; Protein: 1615 AA.

AC AAB59826:

XX 04-APR-2001 (first entry)

DE Protein #3 encoded by Tctd/E gene.

KW Toluene degradation; enzyme; waste degradation; Tctd; Tctd.

XX Thauera aromatica.

XX Xanthomonas maltophilia.

XX Geobacter metallireducens.

XX Acetococcus toluilyticus.

XX WO200072650-A2.

XX 07-DEC-2000.

XX 24-MAY-2000: 2000MO-US14298.

XX 01-JUN-1999: 99US-0323872.

XX (UOH-) UNITV OHIO.

XX Coschigano PW.

XX WPI: 2001-041080/05.

XX N-PSDB: AAF23627.

XX



38 GGGGCTGACACTCGTGCATCCTCGTGGCTGAAGACAAGGCTAT 87  
:::||| ||||::: ::||| ::::~::~:

4 SerAlaThrThrProLysIleProThrCys.....SerArgCysI1 17

88 GAGCGATCGTCCCTATCTGAGCA.....ACATTGGCCAGAAAGAAAGCTT 131

17 eProAnCysProThrThrProCysArgThrThCysIleAlaThrThrIa 34

132 CGAG.....AAGCCAGGAANAAGAGCACTAGCTTGGGGCCA...AAAG 175

34 rArgSerArgProThrThrAlaAlaAlaAlaSerMetAsnThrGlySerArg 50

176 TGATCATGTGAGATGTCACAGGAGGAGTTTG..... 205

51 IleAlaCysArgAlaSerAlaSerProIleSerIleArgGlnThrSerAl 67

206 .....TGAGAGGATTCATCGCCGCCATCCA 233

67 AlaIaCysIleArgSerAlaIaIaThrPArgArgProSerLeuProAlaI 84

234 GTCCAGGCGCACTGATTAGAGACCGCTACCTCTG..... 268

84 hThrAlaSerCysArgProThrThrLysProSerThrPylsThrGlyCysTrp 100

269 ..GCACCTCTCTTGCAGGCCCTGCATGGCC..... 298

101 ArgAlaSerSerSerProLysSerIleSerProLysProArgProThrCys 117

298 ..... 298

117 sArgProSerProGlyThrAlaAlaAlaAlaValSerThrThrSerProArgS 134

299 .....GCAACAGCAAGTGAATAATGCCCGCCAGCGGAG..... 328

134 eThrThrGlyArgArgPArgPArgSerProAlaAlaAlaAlaSerAlaGlyArg 150

329 .....GGGCCAAGATGATGTGCCACAGGCGCCACAGGAAGAGGA...ACGAT 372

151 AlaGlyAlaAlaGlyCysAlaArgSerSerArgGlyThrSerArgProIle 167

373 CAGTCCGCGTTTGAAGCTCA.....GCTGCTATCTCACT 404

167 eArgSerAlaAlaArgProSerCysSerLysSerProThrSerAlaThrIaP 184

405 GGGCCCCCAGA.....TAAAGTCATTG 427

184 hEProProSerProAlaAlaAlaAlaSerArgThrArgCysArgAlaArgAsnSer 200

428 CTCCTCTGAGGA..... 439

201 LeuProSerSerAlaThrArgSerSerAlaThrArgAlaAlaIaThrProAr 217

440 .....TGCTGAAATCTCTCAAC..... 457

217 gArgLysThrProCysGlyArgThrThrArgProProSerSerThrIa 234

458 .....GGTTCAGAGCGCGCAATGACTGATGC...AGTACGGAAGAAGAC 499

234 rGAsnSerSerArgAlaIaThrTrpMetArgTrpAsnSerSerArgTrpAsn 250

500 ACGGGATTCCATCCCGGCTCA.....CTCCCAAGAAGCCCGTGG 537

251 ValArgPArgProSerMetAlaProAlaSerArgAlaProThrAlaLysSe 267

538 AGCATGATGAGAACTCATGACCA.....TCAGTACAGAGCTGG 578

267 rSerArgLysArgThrIleCysSerSerSerProSerAlaIaIaIaProThrP 284

579 AATCTCGAGAACCCCAAGAACCAAGGCGCTCCAGGCTCTTACACGAGAC 628

284 rArgAlaAlaArgThrProAlaIaThrThrProThrPro...SerSerArgGln 299

629 GCC.....AGAGCCACCAAGAGCCCGCCACACCCCTGACA..... 664

300 ProSerGlySerAlaArgProSerProProSerSerSerAlaIleProAr 316

```

XX WPI: 1999-069812/06.
DR N-PsDB; AAW82910.
XX
PT DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
XX assembly of functional glutamate receptor subunits
PS
XX Example 3; Columns 261-268; 203pp; English.
XX
CC The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC receptor subunit (NMDAR). The nucleic acid sequence does not contain the
CC 860 5' most nucleotides, has an additional 11 nucleotides (AAW82891)
CC between nucleotides 1300 and 1301, an additional 24 nucleotides (g.
CC (AAW82890) inserted between nucleotides 23050 and 2351, as set forth in
CC AAW82889. The cDNA sequence is derived from clone NMDA24. The NMDAR
CC ion channels. In addition to being useful for the production of NMDA
CC receptor subunit proteins, the nucleic acids are also useful as probes to
CC identify and isolate nucleic acids encoding related receptor subunits.
CC Functional glutamate receptors can be assembled (homomeric) or from combinations
CC of several subunit proteins of one type (heteromeric). The present invention
CC discloses methods for using such receptor subunits to identify and
CC characterize compounds which affect the function of such receptors, e.g.
CC agonists, antagonists and modulators of glutamate receptor function. The
CC invention also comprises methods for determining whether a protein(s)
CC protein(s) are functional as NMDA receptor subunits.
XX Sequence 1061 AA:
XX
XX Alignment scores:
XX Quality: 174.00 Length: 469
XX Ratio: 0.961 Gaps: 25
XX Percent similarity: 38.593 Percent identity: 23.881
XX
XX alignment block:
XX US-09-775-693-I x AAW87504 ..
XX
Align seg 1/1 to: AAW87504 from: 1 to: 1061
38 GGCGGCTGGACACT.....CG 54
|||||:|||||:
520 AlasertprserProleuglyLeualaIaySerLeuleProleualath 536
||| ||||| |
55 TGATCCTCTGTGCTGAAGACAAGGCTTGAAGTCATTGCCATCT 104
||||| ||| ||| |
536 rAlaserProcys.....ArgAtgrThrProthrGlySerGlyProhmTr 551
||| ||| ||| |
105 GGCCAACATTGGCCAGAAGAACTTGAGG..AACCCAGGAAGAG 151
|| : : : : |||: ||| :
551 rpgrsgcysserSerrtpglYthrgluahlsarAsnTPRtvgln 567
||| : : : : |||: ||| :
152 CACTGAAGCTTGGGCCAAAAGGTTCAATGAGATGTCAGAGGAG 201
||| ||| ||| |||
568 cyseglySerclnglySerAlary.....MetArqatgrHrrArqAlal 582
||| ||| ||| |||
202 TTTTGAGAGAGTTCATCGCCGCATCCAGCCAGCCACATGTATGA 251
: ||| |||: : : |||||
582 aserTrprThrSerThrThrTpgrlualaser..... 593
||| |||: ||| |||
252 GGACGCCTAACCTTCGGGACACCTCTCTTGGCACAGCCCTGCATCCCCGA 301
|||||: ||| |||
594 ..ThrcycystsrPrtrProtprlpolyTrrpocystPrpsrserProgly 609
||| : : : : |||: ||| :
302 AACACAGTGAATAAGCCACGCGAGGAGGAGGACCAAGTATGTGCCACAGG 351
||| : : : : |||: ||| :
610 SerThrTrpSer.....ThrglySerCyalaIathArqySerThrhn 624
||| : : : : |||: ||| :
624 sPrserTrThrSerSer..... 624
||| : : : : |||: ||| :

```

[illegible]

```
seq_name: /SIDS2/gcgdata/geneseq/geneset-1  
seq_documentation_block:  
seq_annotation_block: A087503 standard; Protein; 1212 AA.
```

DNA encoding N-methyl-D-aspartate receptor subunits assembly of functional glutamate receptor subunits

xx	Sequence	1212 AA;	Length:	465
sq			Gaps:	2
	alignment-scores:		Quality:	174.00
	Ratio:	0.961		

Percent Similarity: 38.593 Percent Identity: 23.881

alignment\_block:

US-09-775-693-1 x AAW87503 ..

Align seg 1/1 to: AAW87503 from: 1 to: 1212

```

38 GCGGCTGAGACACT.....CG 54
671 AlasertPserProLeuGlyLeuAlaArgSerLeuProLeuAlaThr 687
55 TGCAATCCCTGCTGCTGAAGACAAAGCATATGATCATGCTATGCTATC 104
687 rAlasertProCys.....ArgArgThrProThrGlySerGlyProThr 702
105 GCGCAACATTTGGCCAAAGAAAGCATGCGAG...AAGCCAGAAAGAG 151
702 rArgSerCysSerSerThrGlyThrGlnArgHisArgSerThrArgGln 718
152 CACTGAAGCTTGGGCGCAAAAGGCTTCTATGAGATGTCAGAGGAG 201
719 CysGlySerGlnGlySerAlaArg.....MetArgArgThrArgAlaAl 733
202 TTGTGAGAGAGTATCATGCGCGGATCCAGTCCAGCGAGCTGTATGA 733
733 aSerThrPserThrThrThrThrProAlaSerSer..... 744
252 GAGCGGCTACCTCTGGGCACTCTCTGCGAGCGCCGATCGCCGGA 301
745 ..ThrCysTrpTrpProThrProGlyTrpProCysTrpSerSerProGly 760
302 AACAAETGAAATTCGCGCGAGGAGGCGCAATGTCGTCAGCGGC 351
761 SerThrTrpSer.....ThrGlySerCysAlaThrArgCysSerThr 775
352 GCCACGAAAGGGAACG...ATCAGTCCGCTTGAAGCTGCTGCT 775
775 sProSerThrPserCysTrpLeuSerAlaGlyAlaSerThrAlaAla 792
398 ..... 792
792 eAlaGlyCysAlaArgAlaSerProAlaHisArgGlyAlaProAlaAla 808
401 CACTGGGCGCCAGATAAAGGTCATTCGCTCCGAGAGATCGCGAATTC 808
809 SerArgProAlaArgProArgProAlaCysSerArgPheCysArgGln 825
451 TACACCGGTTCAAGGCGCGCATGACCTCATGAGCT 825
825 oAlaThrTrpProArgAlaAlaAlaThrProThrAlaProLeuAla 842
492 AAAGCAACAGGGGATTC.....CATCCGCGTCACTCCCA 842
842 rSerArgGlyGlyValAlaAlaValAlaArgProHisArgProProAla 858
527 AGAACCCCTGGAGCATGATGAGAACCTGATCAGATGAGGAGCT 858
859 ArgProArgGlyLeuAlaProAlaHisAlaCysProProThrArgP 875
577 GGAATCCCTGAGAACCCCAAGA.....ACCA 602
875 oGlnSer...ArgAlaProArgAlaGlyAspArgGlnThrGlyValAla 891
603 AGGCGCTCCAGGCTCTACAGAGACCCAGACCCAGCAAG..... 646
891 rArgGlyLeuCysAlaGlyLeuArgSerProArgAlaAlaProArgArg 907
647 .....CCCCCAACCCCTGACATCTCGAGATCGAGTTCAA 684
908 GlyArgProCysProThrSerProGlyCysArgAlaGlnProGlyArg 924
685 AAAGGAGCTCCCTGAGAGTGACCAAGGTCAGATGCGACCCACCC 733
..... 733

```

```

924 gArgGly.....GlyArgCysGlyProGlyThr 934
734 .....AGACCTCTTGAGAGCTTCATGAT 757
934 lArgGlyThrSerArgProPro...SerGlyProCysArgProArgAla 949
758 ..... 950
950 ValThrThrAlaProPheLeuGlnProThrAspProAlaAlaProSer 966
763 AACGAGATGCGCGCAAGCATGCGCTG.....GCCGAT 797
966 rArgSerSerArgSerProArgSerThrArgThrCysArgCysSerVal 983
798 TGACATGCTGGAGAACCGCTCATTTGCAATGAGTCCGAGTATCTAG 847
983 rSerSerThrPro.....GlyAlaArgProCysThr 993
848 AGACCCGACGAGCGACATCTTACCATGCTCATTTAGACATCGAGCC 897
994 ArgPro.....GlyProGlyAlaArgAlaArgValThrLeuProCys 1008
898 TTCACCATGAGACCGGGAAGTGGCGAAATCAACAAGGCTGGGCTTGA 1049
1008 oAlaProThrProArg..... 1013
948 ATTGCTGAGCTGTGTATACGGGTTACGGCTTACGCTGAGTGAAT 997
1014 .....ProSerLeuGlyProAla...ArgCysPro 1022
998 TTGCGCGCCAGTGCATCGCAAGT.....CCGAG 1026
1023 LeuGlyAlaProAlaProProAlaProAlaProThrAlaThrArgPro 1039
1027 GAGCAGTGGAAAGGGAAGTGCAGTGTCCGCTCAAGGCGCAGTGTGA 1076
1039 aGlyAlaThrPheArgArgSerArgCysAla..... 1049
1077 CATCTCGCGCGGAGTCCCACTGCTCTTCAATGAGAGGCTGGTA 1126
1050 .....CysArgSerThrGlyArg..... 1055
1127 GCATGAGCTGAGGAGGATTTAGACCACTGATGCGCGGCTTCATC 1176
1056 .....ProAlaArgArgAlaSerArgGlnGlyProProGlySerThr 1070
1177 AACATCA 1183
1070 rAspSer 1072

seq_name: /SID2/gcdata/geneseq/geneseq/AA2000.DAT: AAB26240
seq_documentation_block:
ID AAB26240 standard; Protein: 1081 AA.
AAB26240;
23-FEB-2001 (first entry)
DE Human N-methyl-D-aspartate receptor subunit NMDAR1 #18.
KW Human; N-methyl-D-aspartate receptor; NMDA; NMDAR1; Ionotropic;
KW glutamate receptor; drug screening; animal model; disease diagnosis;
XX genetic screening.
XX Homo sapiens.
XX OS
XX
XX Key
XX Location/Qualifiers
XX Misc-difference 116
XX /label= unknown
XX /note= "encoded by TGA"
XX Misc-difference 140
XX /label= unknown

```





```
745 ProArgAlaGlyAspArgInThrGlyValAlaArgArgLeuCysAlaG 761
619 TACACGAGAGACCCAGACCCCAAG.....CCCC 650
: ||| |||||: |||:
761 yLeuArgSerProArgAlaAlaProArgArgGlyArgProCysPro 778
651 CAACACCCCTGACATTCGAGATCGATTCAAAAAAGGGCTCCCTGTGA 700
: ||| : : : : : |||
778 hrSerProGlyCysArgAlaAlaGlnProGlyArgGly..... 791
701 AGGTGACCAACGTCAAGATGCGACCCACC.....AG 735
792 .....GlyArgCysGlyProGlyThrAlaGlyGlyThrSerAr 804
736 ACCTCCTTGAGGCTTCATGT..... 757
||||| |||: |||
804 gProPro..SerGlyProCysArgProArgAlaValInThrAlaPro 820
758 .....ACCTGAACGAAGTCGCGGCA 778
820 heLeuGlnProThrAspProAlaAlaProSerSerArgSerSerArgSer 836
779 AGCATGGCGCTG.....GCCGTATGACATCGTGAGAAC 813
: ||| : ||| : ||| |||
837 ProArgSerThrParGThrCysArgCysSerValArgSerSerThrProG 853
814 CGCTTCATTGGAATGAAGTCCGAGATATCTACGAGACCCAGCAGCAC 863
: : : : ||||| |||
853 yGlyArg.....ProCys***ThrArgPro.....GlyP 863
864 CATCCTTTACCATGCTCATTTAGACATCGAGGCTTACCATTGACCGGG 913
: : : : |||||
863 roGlyAlaArgAlaArgValThrLeuProCysProAlaProThrProArg 879
914 AAGTGGCAAAATCAACAAGAGCGCTGGCTGAAATTGCTGAGCTGTG 963
879 ..... 879
964 TATACCGGTTTACGCCCTAGACCTGAGTGAATTGCGCCACTGCAT 1013
||| ||| |||: ||| |||
880 ....ProSerLeuGlyProAla...ArgCysProLeuGlyAlaProAlaP 894
1014 CGCCAGT.....CCGAGAGCGAGTGAAGGA 1042
: : : : ||| : : : |||
894 roProAlaProAlaProThrAlaThrArgProAlaGlyAlaThrParArg 910
1043 AAGTGCAGGTGCCCTCCTCAAGGGCCAGGTACATCCTCGCGCGGAG 1092
: : : |||||
911 ArgSerArgCysAla..... 915
1093 TCCCCACTGTCTCTACAAATGAGAGCTGTGAGCATGACGTGCAGGG 1142
||| ||||| ||| : |||
916 .....CysArgSerThrGlyArg.....ProAlaArgA 925
1143 TGATTATGAGCACTGATGCCAGCGGTTTCATCAACATCA 1183
||| ||||| |||: |||
925 rglAlaSerArgGlnGlyProProGlySerThrAspSer 938
```





```

374 .... ||||| ..... |||
       ProGlyProGlyAlaGlnGlyProAlaGlyProGlyGlySerA 389
727 ACCGACGAGACCTCTGGAGCTCTTCATGTAACCTGACAGACAGTCCGGG 776
       ||| ..... |||
       rgAspProGly.....ProGlyAlaGlnGly 388
777 CAAGCATGGCTGGCCCTATTGACATCGTGAGAGACCG...CTTCATWG 823
       ||| ..... |||
       ProAlaGlyProGlyGlySerArGAspProGlyProGlyAlaGlnG 415
824 GAATGAAGTCCCGAGGTATCTACGAGACCCGAGCAGCACCATTCTTAC 873
       ||| ..... |||
       yProAlaGlyProGlyGlySerArGAspPro..... 425
874 CATGCTCATTTAGACATGAGGCCCTTCACCATGAGCCGGAAGTCCGAA 923
       ||||| ..... |||||
       426 ..... GlyProProGlyAlaGln 431
924 AATCAACAAGGCTGGGCTTGAAATTGCTGAGCTGGTATACCGGTT 973
       ||| ..... |||
       432 GlyProAlaGlyProGly..... 437
974 TAGGCTAGCCCTGAGTGTGAATTTCGCCCACTGCATGCCCAAGTC 1023
       ||| ..... |||
       438 ..... GlySerArGAsp 442
1024 CAGGAGGAGTGGAGGAAAGTCAAGTGTCTGCTTCAGAGCCAGGT 1073
       ||| ..... |||
       442 roGlyProProGlyAlaGlnGly.....ProAlaGlyProGly 454
1074 GTACATCTCGCCGGAGTCCCACTGTCTCTTCATGAGAGCTGG 1123
       ||||| ..... |||
       455 GlySerArGAspProGlyProPro.....G 463
1124 TGAGCATGAACGTGACAGG 1142
       ||||| ..... |||||
       463 yAlaHisGlyProAlaGly 469

```

seq\_name: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:US-09-041-886-23

seq\_documentation\_block:

```

: Sequence 23. Application US/09041886
: Patent No. 6235872
: GENERAL INFORMATION:
: APPLICANT: Bredesen, Dale E.
: APPLICANT: Radizadeh, Sharoz
: TITLE OF INVENTION: Propoptotic Peptides, Dependence
: TITLE OF INVENTION: Polypeptides and Methods of Use
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/041,886
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 2626
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949

```

```

: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1185 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-041-886-23

```

```

alignment_scores:
  Quality: 139.00      Length: 439
  Ratio: 0.755
  Percent Similarity: 41.913      Percent Identity: 26.651

```

alignment\_block:

US-09-775-693-1 x US-09-041-886-23

Align seg 1/1 to: US-09-041-886-23 from: 1 to: 1185

```

21 GGTTCGGCTACAGTGGCGGCTTGAGACCTCGTGCATCTCGTGGC 70
   ||| ||| ||| ||||| ||| |||
319 GlyLeuGlyAlaGlnProLeuProGlyHisLeuProSerPro...TyrAl 334
   ||| ||| ||| ||| ||| ||| |||
71 TGAAGAACAGGCTATGACGTGATTCCT.....ATCTGGCC 108
   ||| ||| ||| ||| ||| ||| |||
109 AACATGGCCGACGAAGACATTCGAGAACCCAGGAAGACGACTGA 158
   ||| ||| ||| ||| ||| ||| |||
350 omrLeuAlaProSerProHisSerLeuProAlaSerSerAlaP 367
   ||| ||| ||| ||| ||| ||| |||
159 GCTTGGGCGCAAAAG.....TGTCATTGAGATGTCA 193
   ||| ||| ||| ||| ||| ||| |||
367 roAlaProPomelargPheProTyrSerSerSerSerSerAla 383
   ||| ||| ||| ||| ||| ||| |||
194 GCAAGGAGTTTGTGAGAGATTCTGCGCGCCATCCAGTCCAGCGCA 243
   ||| ||| ||| ||| ||| ||| |||
384 AlaAlaSerSerSerSerSerSerSerSerAlaSerProPhePr 400
   ||| ||| ||| ||| ||| ||| |||
244 CTGTATGAGACGCTACTCTCTGGGACCTCTCTGCGAGGCGCTGAT 293
   ||| ||| ||| ||| ||| ||| |||
400 oAlaSerGlnAlaLeuProSerTyrProHisSerPheProPheProHis 417
   ||| ||| ||| ||| ||| ||| |||
294 CGCCCGCAACAGTGAATCGCCGAGGAGGAGGCGCAAGTATG... 340
   ||| ||| ||| ||| ||| ||| |||
417 erLeuSerValSerAsnGlnProProTyrTyrThrGlnProSerLeuPro 433
   ||| ||| ||| ||| ||| ||| |||
341 .....TGTCACGCGCGCCACAGAAAGGGGAAAC 369
   ||| ||| ||| ||| ||| ||| |||
434 SerGlnAlaValTyrSerGlnGlyProProProProProProTyrGlyPr 450
   ||| ||| ||| ||| ||| ||| |||
370 GATCAGGTCCGCTTGGAGCTGAGTGTACTGACTGACCCCGCAGATAA 419
   ||| ||| ||| ||| ||| ||| |||
450 glu.....LeuAlaAsnSerAsnAlaHis..... 458
   ||| ||| ||| ||| ||| ||| |||
420 GGTCAATGCTCCCTGAGAGATGCTGAATTCATACACGGTTCAAGGCC 469
   ||| ||| ||| ||| ||| ||| |||
459 .....ProGlyProPheProProSerThrGlyAlaGlnSerThr 471
   ||| ||| ||| ||| ||| ||| |||
470 GCAATGACCTGATGAGTAC..... 490
   ||| ||| ||| ||| ||| ||| |||
472 AlaHisProProValSerThrHisHisHisHisHisGlnGlnGlnG 488
   ||| ||| ||| ||| ||| ||| |||
491 .....CAAGCAACAGCGGA...T 506
   ||| ||| ||| ||| ||| ||| |||
488 nGlnGlnGlnGlnGlnGlnGlnGlnGlnHisHisGlnGlnGlnG 505
   ||| ||| ||| ||| ||| ||| |||
507 TCCATCCGGGTCACTCCAGAACCCGAGAGCATGATGAAGAACSTCA 556
   ||| ||| ||| ||| ||| ||| |||
505 roProProGlyAlaPheProHisProLeuGlnGlnGlySerSerHis 521
   ||| ||| ||| ||| ||| ||| |||
557 TGACAT.....CAGCTAGAGGCTTGA 579
   ||| ||| ||| ||| ||| ||| |||

```



```

153 rgleu1leproarg...ProhaleucyProtrpserCysleuargArg 168
449 TCTACACCGGTTCAAGGCGCGCATGACCTGATGAGTACGCAAGC... 496
169 AsnProThrSerProArg1SerSerThrProThrAlaThrAlaSerAr 185
497 ...AACACGGATTCACATCCCGGTCCATCCCAAGAACCGGTGAGCAT 542
185 gleuGlnProSerThrProSerSerSerThrProSerleuSerGly... 200
543 GATGAGACCTCAGACACATCAGCTACGAGGCTGGAATCCTGAGAAC 592
201 ...SerThrAlaProGlyThrThrSerThrProSerCysSerPro 214
593 CCAAGAACCAAGCGCTCCAGGCTCTACAGAGAACGCC... 631
215 ProleuSerSerArgArgInleucySerArgThrProProSerPheTh 231
632 ...AGGACCCAGCCCAAGCCCGCCAGACCCCTGACATTCGAGATGCA 677
231 rglYArgThrThrThrAlaAlaSerProleuSer... 243
678 GTTCAAAAAAGGGTCCCTGTGAAGTGACCAACGTCAGATGAGACCA 727
244 ...CysAlaArgleuPro 249
728 CCC...ACCAGACTCTCT...TGGAGCTCTTCATG 756
250 ProleuSerSerArgProProThrSerAlaGlyCysThrGlnleuSer 266
757 TACCTGAACAG... 769
266 sThrProAlaSerProProHisCysProAlaGlnGlySerSerGly 283
770 ...TCGCGGCGACCATGCGC 787
283 lYThrPheSerleuProGlyGlyPheleuSerSerProAlaCysGly 299
788 TGGCCCGATTGACATCGTGGAGAACCGCTTATGATGATGATGCCGA 837
300 Trp...ValAlaSerTrpProleuSerMetleuArgAlaGlyThrTr 314
838 GGTATCTAGAGAGCC...CACGAGGACACATCC... 868
314 pThrThrAlaProProThrGlyGlySerSerLysAspProSerSerTr 331
868 ... 868
331 euleuGlyThrleuGlyPhePheSerIleleuSerValSerCysGlyAsn 347
869 TTTACCATGCTCTATTAGACATCGAGCGCTTCACCATGAGCGGG... 913
348 TrpSerGlnleuArgAlaAlaSerThrProSerleuSerThrGlyValSe 364
914 ...AAGTGGCAAAATCAAAACAGGCT 938
364 rleuSerGlnArgPheSerSerPheArgCysleuGlnSerThrMetSerP 381
939 GGGCTTGAATTTGCTGAGCTGTATACCGGTTTACGGCTACGCCCTG 988
381 heser...ThrSerCysleuThrValleuGly... 390
989 AGTGTGATTTGTCCGCACTGATCGACCAAGTCCAGAGGAGGATGGA 1038
391 ...TrpThrSerAlaSerProAsnTrpAspTrp... 400
1039 GGGAAAGTGCAGGTGCTCGTCTCAAGGCC...AGGTGA 1076
401 ...AlaLeuSerArgAlaSerleuLeuSerCysT 412
1077 CATCTCGGCC 1087
412 hrAlaSerSer 415

```

```

seq_name: /cgn2.6/plodata/2/laa/6B_COMB.pep:US-09-413-814-78
seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bioecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichembach, Hans
; TITLE OR INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: patentin Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-413-814-78

alignment_scores:
Quality: 134.00 Length: 433
Ratio: 0.784 Gaps: 24
Percent Similarity: 39.492 Percent Identity: 25.173

alignment_block:
US-09-775-693-1 x US-09-413-814-78
Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

107 CCACATTGGCCAGAGAGACTTCGAGAGCCAGAGAGAGGACTG 156
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
358 ProSerArgArgArgSerProAlaArgThrProGlyProArg... 372
157 AAGCTTGGGGCCA...AAAGTGTTCATTGAGATGTCAGGAGGACTT 203
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
373 ...SerProArgArgArgProSerProSerAlaArgSerProAlaI 387
204 TGTGAGAGAGTTCATCTGGCGGCGCATCCAGT...CCAGCGCATGT 247
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
387 lerrpGlnArgProProArgArgProArgAspArgArgProAlaArgleu 403
248 ATGAGAGACCGCTACCTCCGCGCACCTCTTGCAGGC...CTGC 291
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
404 leuArgProAlaArgArgleuAlaArgAlaArgAspProLysValProAs 420
292 ATGCGCCCGCAAAAGTGAAGTGAATGCGCAGC... 322
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
420 pProAlaAlaAlaArgArgAlaProAlaAlaGlyAlaLeuProGlyA 437
323 ...GGAGGGGGCCCAAGTATGTGTCCACGCGCCACAGAGAGGGA 367
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
437 laaspolyArgGly...AlaArgGlyProGlyArgArgArg 449
368 AGCATAGAGTCCGGTTTGAAGCTCAGCTGCTACATCGAGCGCCCGCCAGATA 417
111: 111: 111: 111: 111: 111: 111: 111: 111: 111:
450 AlaProArgArgGly...AlaAlaAlaArg.ValProProAsp 463
418 AAGTATGTCCTCGAGAGATGCGTGAATTCATCAACCGCTTCAAGGG 467

```

```

463  ....
466  CCGCATGACCTGATGATGACGCAAGCAGCAGGATTCCTATCCCGG 517
476  ArgAlaAlaProAlaValAlaArgSerAlaArgAlaGlnProG1 492
518  TCATCCCAAGAACCCGTGACATGATGACACCTCATGCATCAGC 567
492  yValGln.....HsProv 497
568  TCGAGGCTGGATCTTGGAACCCCAAGCAGCCCTCAGGCT 617
497  AlArgG1AlaArgAspGly.....ProAlaArgArg... 508
618  CTACACGAAGACCCAGACCCCAAGCCCAACACCCCTGACATTC 667
509  ....ArgAlaProProGlnPro.....Va 515
668  TCGAGATGAGTTCAAAAAGGGTCCCTGTGAAGTGACCAAGCTCAG 717
515  lArgAspArgAlaAlaProArgG1AlaAlaAspAspValGlnArgGln 532
718  GATGACACCCACCCAGAC.....CTCCTTGAGCTTTCATGTACT 761
532  lYArgArgAlaValProG1YArgArgAlaArgG1ProG1AlaValPro 548
762  GAA...CGAAGTCGGGGCAAGCATGGCTGGGCGG..... 794
549  AspValArgProArgAlaProArgG1YArgG1ProArgArgG1YLe 565
795  ....TATTGACATGCTGAGAACCGCTTCATGGAATGAGTCCGAGCT 840
565  uG1YAlaArgProArgG1YlurProArgAlaValAlaProG1lurPro 582
841  ATCAGACAGACCCAGCAGCAGCAGCAGTTCATGCTGATGAGCT 890
582  lAlaAlaProProAlaG1AlaH1sProG1AlaProArgArgAlaArgH1s 598
891  CGAGGC.....CT 898
599  ArgAlaArgArgAlaProArgArgValAlaArgG1eUValG1YArgArgLe 615
899  TCACCATGAGACCGGAGAGTCCCAAAATCAAAAGCGCTGGCTGAAA 948
615  uArgArgArgAlaArgArgAlaLeuArgArgLeuArgAlaG1urProAla 632
949  TTGCTGAGCTGTGTATACCGGTTAGGCGCTAGCCTGAGTGAATT 998
632  hProAlaAlaG1Y..... 636
999  TGTCCGACCATGATCCCAAGTCCCAAGCAGC.....AGTGAAG 1039
637  ....AlaProG1YAlaValAlaArgArgLeuArg 646
1040  GGAAGATCAGGTGTCCT...CCTCAAGGCGCAGGTGTCAT..... 1079
646  gSerProAlaG1YAlaAlaValAlaArgG1YProG1YAlaAlaPro 663
1080  ....CCTGCGCGGCA 1091
663  AlLeuAspAspG1uAlaLeuG1YArgAlaAlaProG1YProAlaG1Y 679
1092  GTCCCACTGTC.....TCTCTCAATGAGACCTGGTAGCA 1129
680  ArgProArgAlaG1YAlaAlaAspValAlaArgG1YAlaAlaG1ur 696
1130  TGAACGTGACAGGTGATTAAGCAACTGATGCGCGGTTCAATCAAC 1179
696  oArgArgArgG1Y.....ArgAlaH1s.... 703
1180  ATCAATTCCTCAGGCTGAAGGATATCTGCTTCACAGCAGCAAGT 1226

```

```

704  .....ProProAspGlnG1Y 708
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-483-533-41
seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Rolzman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method for Treating Tumorigenic
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-533-41

alignment_scores:
Quality: 131.50 Length: 343
Percent Similarity: 37.026 Percent Identity: 24.490

alignment_block:
US-09-775-693-1 x US-08-483-533-41 ..
Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

206 TGAAGAGTTTCATCTGCGCGGCATCCAGCGCACTGTATGAGAC 255
111 ..:111 11111111
82 TrpProAspSerProProG1uSerAlaProG1u.....AlaAr 95
256 CGCTACCTCTGCGCACCTCTCTTGCAAGCCCTGCAATCGCCGCAACA 305
11111111 11111111
95 gProTrpAlaAlaAlaProArgProProG1YProH1sArgProAla 110
111 ..:111 11111111
306 AGTGAATTCGCCAGCGGAGGCGCAAGATGTGTCTCCAGCGCCA 355
111 ..:111 11111111

```



```

111 ..TrrpalaaraglyAlaGlyLeuThrPro.....ProThrProPro 123
356 CAGGAAAGGGGAGACGATCAGGTCCGGTTTGAGCTCAGCTGCTACTCACTG 405
124 ArgAla..... 125
406 GCGCCCGCATTAAGTTCATGTCCTCCCTGAGAGATGCCCTGATTCTACAA 455
126 .ProSerAlaPheArGAlaSerProSerAlaCysAlaSer..... 139
456 CCGGTTCAGAGGCCCGCATGACCTGATGAGTACGCAAGAACACAGCGGA 505
139 ..... 139
506 TTCCCATCCCGGTACTCTCCCAAGACCGGTGAGCATGATGAGAACCTC 555
140 .....ProArGSerThrTrpArGAlaCysAla 148
556 ATGCACATCAGCTACGAGGTGGAATCTGAGAACCCCAAGAACCAAGC 605
148 acysAla.....ArgAlaGlyArGAlaArG 158
606 GCCTCCAGGTCTCTACAGAGACCCAGACCCCAAGCCCGCAACA 655
158 rg.....SerProArGProProArGProProArGProProArG 171
656 CCGCTGACATTCCTCGATCGAGTTCAGAAAGGGGTCCGTGAGGCTG 705
172 ProProArGProProArGlyCysAlaSerArG..... 182
706 ACCAAGTCAGAGATGACCA.....CCCAAGACCTCTT 743
183 .ProThrSerGlyCysAlaThrTrpSerGlyProArGProProAla 199
744 GGAGCTCTTCAGTA..... 758
199 rPArGAlaAlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySe 215
759 .....CCTGACGAGATCGCGGACAGATGCGCTG 789
215 rGlyAlaGlyTrpArGArGProArGArGSerSerGlyArGAlaTrp...G 231
790 GGCCGTATTGACATCGTGAGAACCGCTTCATGGAATGAGTCCCGAGC 839
231 lyPro.....ArgProValProGlyProTrpProAlaGluProAla 244
840 TATTCAGAGACCCAGACGACCATCTTACCATGCTCATTTAGACA 889
245 ..ArgArGThrArGSerSerAlaThrPro...GluAlaAlaTrpValPh 259
890 TCGAGGCTTCACCATGACCGGAGATGCGCAAAATCAACAGAGCTG 939
259 eArGly.....AlaProGlySerSerAlaProSerArGSerProG 273
940 GGCCTGAAATTTGCTGAGCTGTATACCGCTTACGGCCTACGCCCTGA 989
273 lArGArGTrpGlnGluProArGlyIleTrpThrLeuGlyAlaSerProPro 289
990 GGTGAAATTTGCGCCACTGCATGCGCAAGTCCAGAGCGAGGAGG 1039
290 SerGlnGlyGlyProProArGlyArGAspTrpPro...IleGlyGlyAr 305
1040 GGAAGTGCAGGTGTCGT..... 1058
305 gGlnArGlyGlyAlaArGProThrSerValArGArGValPheGlyAla 322
1059 .....CCTCAAGGGCCAGGTGATACATCTCTCG 1085
322 rPProIleGlyArGAlaGluProProAsnGlyProGly...ArgProLeu 337
1086 CCGGAGTCCCGCTGCTCTACAA 1112
338 ProGlyMetValIleLeysAsnLeuGln 346

```

seq\_name: /cgn2.6/ptodata/2/1aa/PCTUS.COMB.pep:PCT-US91-06532-3

seq\_documentation\_block:

: Sequence 3, Application PC/TUS9106532

: GENERAL INFORMATION:

: APPLICANT: Roizman, Bernard

: TITLE OF INVENTION: Recombinant Herpes Simplex Viruses

: TITLE OF INVENTION: Vaccines and Methods

: NUMBER OF SEQUENCES: 8

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

: STREET: Two First National Plaza Suite 2100

: CITY: Chicago

: STATE: Illinois

: COUNTRY: USA

: ZIP: 60603

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US91/06532

: FILING DATE: 19910910

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Gruber, Lewis S.

: REGISTRATION NUMBER: 27373/8235

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 312/346-5750

: TELEFAX: 312/984-9740

: TELEX: 25-3856

: INFORMATION FOR SEQ. ID NO: 3:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 355 amino acids

: TYPE: AMINO ACID

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: PCT-US91-06532-3

alignment\_scores:

Quality: 131.50 Length: 343

Ratio: 1.035 Gaps: 19

Percent Similarity: 37.026 Percent Identity: 24.490

alignment\_block:

us-09-775-693-1 x PCT-US91-06532-3 ..

Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

```

206 TCGAGAGTTCATCTGGCGGCGCATCCAGTCCAGCGACATGTATGAGAC 255
82 TrpProAspSerProProProGluSerAlaProGlu.....AlaAr 95
256 GCGTACCTCTGCGACACTCTCTTCCAGAGGCTTCATTCGCCGCAACA 305
95 gProThrAlaAlaAlaProArGProProGlyProHisArGProAla.... 110
306 AGTGGAAATCGCCCGAGCGGAGGCGCAAGTATGTGTCCAGCGCGCCA 355
111 ..TrrpalaaraglyAlaGlyLeuThrPro.....ProThrProPro 123
356 CAGGAAAGGGGAGACATCAGGTCCGGTTTGAGCTCAGCTGCTACTCACTG 405
124 ArgAla..... 125
406 GCGCCCGCATTAAGTTCATGTCCTCCCTGAGAGATGCCCTGATTCTACAA 455
126 .ProSerAlaPheArGAlaSerProSerAlaCysAlaSer..... 139

```

```

456 CCGGTTCAGAGCGCCGCAATGACCTGATGAGTACGCAAGCAACACGCGGA 505
139 .....
506 TTCCCATCCCGGTCACCTCCCAAGAACCCGTGAGCATGATGAGACCTC 555
140 .....
556 ATGCACATCAGCTACGAGCGCTGGATCTTGAGAACCCCAAGCAACGAGC 605
148 acysaspala.....ArgAlaGlyArgGlyArg 158
606 GCGTCCAGCTGCTCTACAGAACCAAGCCAGCAAGCCCAAGCCCAACCA 655
158 rg.....SerProProArgProArgProProArgProProArg 171
656 CCCCTGACATTCGAGATGAGATGAGTCAAAAAGGGTCCCTGTGAAGTG 705
172 ProProArgProProArgGlyCysAlaSerArg..... 182
706 ACCAGCTCAGAGATGCGACCA.....CCACACAGCTCCTT 743
183 ..ProHisSerGlyCysAlaThrThrPrrpSerGlyProArgProProAla 199
744 GGAGCTCTTCATGTA..... 758
199 rprarglaala..AlaArgGlyProAlaSerGlyProThrGlyLeuGlyse 215
759 .....CCTGACAGAAAGTCGCGGCAAGCATGAGCTG 789
215 rglYAlaGlyTrpArgArgProArgArgSerSerGlyArgAlaTrp...G 231
790 GCGCGATTGACATCGTGGAGAACCGCTTCATGGAATGAAGTCCCGAG 839
231 lYPro.....ArgProValProGlyProTrpProAlaGluProAla 244
840 TATTCAGAGACCCAGCAGACGACACATCCTTACATGCTCATTTAGACA 889
245 ..ArgArgThrArgSerAsnValThrPro...GluAlaAlaTrpValah 259
890 TCGAGCCCTTCACCATGAGACCGGAGAGTCCGCAAAATCAACAGAGCCGTG 939
259 eaGlyGly.....AlaProGlySerSerAlaProSerArgSerProG 273
940 GGCTTGAATTGCTGAGCTGATACCGGTTACGGCTTACGGCTCCTCA 989
273 luaArgArgTrpGlnGluProArgGlyThrLeuGlyAlaSerProPro 289
990 GTGTGAATTGTCCGCGCACTGCATGCCCAAGTCCAGAGCGAGTGAAG 1039
290 SerGlnGlyGlyProProArgGlyArgAspTrpPro...IleGlyGlyArg 305
1040 GGAAGTGCAGGTGTCCT..... 1058
305 gslnaArgGlyGlyAlaArgProThrSerValArgArgValPheGlyAla 322
1059 .....CCTCAAGGCGCAGGTGTACATCCTCGG 1085
322 rgrProIleGlyArgGluLeuProProAsnGlyProGly...ArgProLeu 337
1086 CCGGAGATCCCACTGCTCTCTACAA 1112
338 ProGlyMetValIleLysAsnLeuGln 346
seq_name: /cgn2_6/ptodata/2/lae/5a_COMB.pep:US-08-642-255-72
seq_documentation_block:
; Sequence 72, Application: US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.

```

```

TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALDRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A5556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 484-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-72

alignment_scores:
Quality: 130.00 Length: 397
Ratio: 0.818 Gaps: 27
Percent Similarity: 40.050 Percent Identity: 28.463

alignment_block:
US-09-775-693-1 x US-08-642-255-72 ..
Align seg 1/1 to: US-08-642-255-72 from: 1 to: 1065
105 GCCCAACATTGGCCGAGAAAGACTTCGAGAAAGCCAGAAAGAGCAC 154
277 GLYAlaHISgLYProAlaGlyPro..LysGlyAlaHISgLYProAla.. 291
155 TGAAGCTTGGGCCCCAAAGAGGTTCATGAGATGT.....C 192
292 .....GlyProLysGlyAlaHISgLYProAlaGlyProLysGlyAla 305
193 AGCAGGAGTTTGGAGAGTTTCATGTGCGCCGCTTCAGTCCAGCC 242
305 lAGInGlyProAlaGlyProGlyGlySerArgGlyAspProGlyProPro 321
243 ACTGTATGAGACCGCTACCTCTGGGCA.....CCTCTCT 279
322 GLYAlaGInGlyProAlaGlyProGlyGlySerArgGlyAspProGlyPro 338
280 GCCAGGCCCTGCATCGCCGCAAAACAAGTGAATGCCCGCAGGAGAG 329
338 oProGlyAlaGInGlyProAlaGlyProGlyGlySerArgGlyAspProG 355
330 GGCCA.....AGTATGTCTCCAGGCGCCA...CAGAAAGGAGAACG 370
355 lYProProGlyAlaGInGlyProAlaGlyProGlyGlySerArgGlyAsp 371
371 ATCAGTTCGCTTGTGAGCTCAGCTGCTACCTGCTGCGCCCGCCAGATMAAG 420

```

```

Sequence 7, Application US/07945283
Patent No. 5352596

GENERAL INFORMATION:
APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudotabes Virus Deletion Mutants
TITLE OF INVENTION: Involving The EPO and LTR Genes
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext.513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1958 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-945-283-2

Alignment_scores:
Quality: 128.00 Length: 413
Ratio: 0.877 Gaps: 20
Percent Similarity: 35.351 Percent Identity: 21.792

Alignment_block:
US-09-775-693-1 x US-07-945-283-2 ..

Align seg 1/1 to: US-07-945-283-2 from: 1 to: 1958

221 GCGCGGCATCGACGATCGAGCCGACGTGATGAGACGCGTACTCTGGGC 270
||||| ||||||| :|||
297 GlycerolProteinSerHis..... 304

271 ACCTCTTGCACGAGCCCTGCATGCGCCGCAACAGTGAATCGGCCA 320
||||| ||||||| :|||
305 ProglyProthrProglyProserGlyProserThrSerHisGlnAspPro 321

321 GCGGAGCGGCGCCAAAGTGTGTCTCCACAGCGGCGCCACAGAAAGGAAAG 370
||||| ||||| :|||
321 ro..... GlyProProthrProserThrSer HisHisHisHisHis 335
|||||
371 ATCAGGTGGCGTTTGAAGTCAAGTGTCTACTGACGCGCCGCAAGTAAAG 420
||||| ||||| :|||
335 scGlnGlyProPro..... ThrSerPro..... 342

421 GTCATTTGCTCCCTGGAGAGATGCGTGAATTTCACAACGG...GTTCAAAGG 467
||||| ||||| :|||
343 .....ArgProserThrSer 348

468 CCGCAATACCTCATGATGAGTACGCAACAGCAACAGCGAGTATCCGCGG 517
||||| ||||| :|||
349 HisGlnAspProProGlyGlyGlyGlyProProserGlnGlnThrHisHis 365

```



```

247 TATG.....AGAGCCGCTACTCTCTGGG 269
249 sSerGlyHisProValAsnProValLeuProAlaThrMetLysAlaTrp. 265
270 CACCTCTCTTGGCAGGACCTGATCCGCCGGAACAACTGGAATTCGCC 319
266 .....MetCysProCysGlyThrTrpLysProPro 275
320 AGCGGGAGGGGCCCAAGTATGTGTCCACAGCGGCCGACAGGAAGGGGAC 369
276 Ser.....ThrsSerArgProThrSerProGln..... 285
370 GATCAGGTCCGGTTGAGCTCAGCTACTGCTAGCCGCCCCACATATAA 419
286 .....AlaThrSerSerTrpThrSerLeuGlnL 295
420 GGTCAATTGCTCCCTGGAGATGCTGATTTACAAACCGGTTCAGAGGCC 469
295 euGlyGlnLeuProGlnGlyCysAlaGlyCysAlaAlaAlaProGlnLeu 311
470 GCATGACCTGATGAGTACGAGCAACAGCAGGAGTCCATCCCGGTC 519
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerThrLeuAspProGln 328
520 ACTCCCAAGACCCGTTGGAGCATGATGAGAACCTCATGCAATCAGCTA 569
328 YLysProGln.....MetLysSerProThrAsnTrpT 340
570 CGAGGCTGGAAATCTCTGAGAACCCCAAGACCGCTCTGAGTCTCT 619
340 hr.ProHisValProAlaGlnGlyProGlnAlaSerArgProProLysLe 356
620 ACAGAGAGCCCAAGA.....CCGAGCCAAAGCCCAACAC 657
356 whis.....ProGlyLeuArgAlaAlaProGlnGlnLPro.....P 369
658 CCTGACATTCCTGAGATCGAATTCAAAAAGGGTCCCTGGAAGGTG 707
369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GlnGlnAsp 381
708 CAAGCTCAAGA.....TGGCACCCACCCAGACCTCTTGAGCTTCA 754
382 GlnProlLeuGlnGlnTrpHisLeuArgSerGlnTrpGlnGly...LeuH 397
755 T.....GTACCTGAACGAATCGCGGCGCAACCATGCG 786
397 scYstrLeuSerProGlyLeuArgAlaGlyThrAlaSerGlyAlaLeuTrpA 414
787 GTGG.....GCCGATTCGATCGTGGAGAACCGCTTCATG 823
414 rSerGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
824 GAATGAAGTCCGAGGATATCTAGAGACCCAGCAGGACCATCTTAC 873
428 .....SerLeuArgProAlaGlyGlyProSer..... 436
874 CATGCTCATTTAGACATCGAGGCTTCACCATGACCGGGAAGTGGCA 923
437 .....TrpThrGlyArg..... 440
924 AATCAACAAGGCTGGCTGGAATTTG..... 952
441 .....GlyGlnTrpLeuGlnValLeuGlyGlyGlyPheSerTrpLeu 454
953 .....CTGAGCTGGTATACCGGTTAC..... 976
455 AspLeuSerLeuCysValGlyLeuTrpProCysProSerTrpAlaCysTr 471
977 .....GGCTAGCCCTGAGCTGTGAAT...TGTCCG 1004
471 pHisValHisLeuLeuProGlyGlnAspLysAlaLeuAsnArgMetLeuA 488

```

```

1005 CCATCTGCGCCAGTCCAGAGCGAGTGCAGGAAAGTCAAGTGT 1054
488 lArThrLeuGlnProGlnLProProGlyGlnGlnAlaGlyGlnAlaAla 504
1055 CCGTCTCAAGGGCCAGGTGTACATCC.....TCGGCCGGAGTCC 1095
505 lLeSerArgHisLeuLysProProAspProLleProSerProProThrPr 521
1096 CCATCTCTCTCTACATGAGGAGCTGG 1123
521 ocYsserLeuSerGlyLeuGlnValTrp 530
seq_name: /cgn2_6/plodata/2/iae/6a_COMB.pep:US-09-190-476B-2

```

```

seq_documentation_block:
/ Sequence 2, Application US/09190476B
/ Patent No. 6025204
/ GENERAL INFORMATION:
/ APPLICANT: Fartell, Catherine L.
/ APPLICANT: Martin, Francis R.
/ APPLICANT: Yabkowitz, Rachel
/ TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
/ TITLE OF INVENTION: FACTOR
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Amgen Inc.
/ STREET: 1840 De Havilland Drive
/ CITY: Thousand Oaks
/ STATE: California
/ COUNTRY: USA
/ ZIP: 91320-1789
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/190,476B
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/735,041
/ FILING DATE: 22-OCT-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Marz, Richard J.
/ REGISTRATION NUMBER: 27,657
/ REFERENCE/DOCKET NUMBER: A-414
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 805,447,4112
/ TELEFAX: 805,499,6751
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 539 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-190-476B-2

alignment_scores:
quality: 127.00 Length: 410
Ratio: 0.760 Gaps: 26
Percent Similarity: 40.732 Percent Identity: 25.366

alignment_block:
US-09-775-693-1 x US-09-190-476B-2 ..
Align seg 1/1 to: US-09-190-476B-2 from: 1 to: 539
98 CCTATCTGCCCAACATTGGCC.....AGAGGAAGACTTTCGAG 135
111 111 111111
185 ProSerHisProArgLeuAlaAlaLeuLeuArgLeuProArgLeuArg 201

```

```

136 GAAGCCAGAGAGGACCTGAGCTTGGGCCCCAAAAGGTCTTCA 185
    ||||| |||
201 GlyProLeuArgAlaAspAlaAlaSerMetHisArgValGly...s 217
186 GGATGTCAGCAGGAGTTTGTGAGAGTTCATCTGGCCG... 226
    ::::: |||||
217 ePrAsnAlaAlaValLeuTrpLeuTrpSerSerHisAsnArgAlaAsn 233
227 .....
234 AlaArgLeuGlnValProProAlaArgThrProSerSerPro...ArgCys 249
247 TATG.....AGACCGCTTACCCTCTGG 269
249 sSerGlnHisProValAsnPheValLeuProAlaThrMetAsnAlaTrp. 265
270 CACCTCTCTTGGCAGGCCCTGCATCGCCGCAACAGTGGAAATCGCC 319
266 .....MetCysProCysGlyThrTrpLysProPro 275
320 AGCGGAGGCGGCCAAGTATGTGTCCACGGCGCCACAGAGAAAGGGAC 369
    ||| ||| |||||
276 Ser.....ThrsSerArgProThrSerProGln..... 285
370 GATCAGTCCGGTTGAGCTCAGCTGACTGACTGCGCCCGCCAGATAAA 419
286 .....AlaThrSerSerTrpThrSerLeuGln 295
420 GGTCACTGCTCCCTGGAGATGCCGAATTCATACACCGGTTCAGGGCC 469
    ::| ||||| |||||
295 euGlyGlnLeuProGlnGlyCysAlaGlyCysAlaAlaAlaProGlnLeu 311
470 GCATGACCTCATGAGTACGCAAGACACAGGAGTTCATCCCGCTG 519
    ||||| ::| |||||
312 AlaMetGlyAlaLeuGlnLeuGlnSerArgAsnSerThrLeuAsnProGln 328
520 ACTCCAGAGACCCGTGAGATGATGAGAACCTCATGCACTACAGCTA 569
    : |||: ||| ||| |||
328 LysProGlu.....MetMetLysSerProThrAsnThrT 340
570 CGAGCGTGGAAATCTGAGAACCCCAAGACACAGCGCTCCAGGCTCT 619
    ||| ||||| |||||
340 hr.ProHisValProAlaGlnGlyProGlnAlaSerArgProLysLe 356
620 ACACGAGACCCAGGA.....CCGACGCCAAAGCCCCAACAGCC 657
    ||||| |||||
356 uHis.....ProGlyLeuArgAlaAlaProGlnGlnPro.....P 369
658 CCGACACTTCTGAGATCGATTCAAAAAGGGTCCCTGTGAGAGGTGAC 707
    ||| ||| |||||
369 roGlnHisMetAlaAsp...ValGlnArgAsn.....GlnGlnAsp 381
708 CAACGTCGAAGA...TGGCACCAACCCACAGACTCTTGGAGCTCTTCA 754
    ::| ||||| |||||
382 GluProLeuGlnGlnTrpHisLeuArgSerGlnTrpGlnGly...LeuHis 397
755 T.....GTACTGAACGAGAGTCGCGGAGCAGAGAGCC 786
397 sCysTrpLeuSerProGlyLeuArgArgThrAlaSerGlyAlaLeuTrpA 414
787 GTGG.....GCCGATTCAGACTCGTGGAGAACCGCTTCATG 823
    ||| ||||| |||||
414 rGserGlyAlaTrpAlaAlaAlaProSerSerTrpSerThr..... 427
824 GAATGAAGTCCGAGTATCTACGAGACCCAGAGCGACGACATCTTAC 873
    ||| |||||
428 .....SerLeuArgProAlaGlyLysProSer..... 436
874 CATGCTCATTTAGACATGAGGCTTCACCATGACCGGGAATGCGCAA 923
437 .....TrpThrGlyArg..... 440

```

```

924 AATCAACAGAGCCCTGGCTTGAATTTG..... 952
441 .....GlyGlnTrpLeuGlnValLeuGlnGlyGlyPheSerTrpLeu 454
953 .....CTGAGCTGTGTATACCGGTTTAC..... 976
    ::||| |||||
455 AspHisSerLeuGlyValGlyLeuTrpProCysProSerTrpAlaCysTr 471
977 .....GGCTAGCCCTGAGTGTGAAT...TTGTCCG 1004
471 pHISValHisLeuLeuProGlyGlnAspLysAlaLeuAsnArgMetLeuA 488
1005 CCACGTCATCGCCCAAGTCCACGAGAGCGAGTGGAGAGGAAGTCAAGTGT 1054
    ||||| ::||| |||||
488 lathrLeuGlnProGlnProProGlyGlnGlnAlaGlyGlyAla 504
1055 CCGTCTCTAAGGCGCAGGTGTACATCC.....TCGGCCGGAGATCC 1095
    ||| ::| |||
505 lIeSerArgHisLeuLysProProAspProIleProSerProProThrPrt 521
1096 CCACGTGTCTCTACATAGAGAGCTGG 1123
521 ocysSerLeuSerGlyLeuGlnValTrp 530

```

seq\_name: /cgn2\_6/prodata/2/laa/6A\_COMB.pep:us-09-190-889A-2

```

seq_documentation_block:
: Sequence 2, Application US/09190889A
: Patent No. 6075008
: GENERAL INFORMATION:
: APPLICANT: Fartrell, Catherine L.
: APPLICANT: Martin, Francis H.
: APPLICANT: Yabkowitz, Rachel
: TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Amgen Inc.
: STREET: 1840 De Havilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: USA
: ZIP: 91320-1789
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/190,889A
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/735,041
: FILING DATE: 22-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Mazza, Richard J.
: REGISTRATION NUMBER: 27,657
: REFERENCE/DOCKET NUMBER: A-414
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 805.447.4112
: TELEFAX: 805.449.6751
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 539 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-190-889A-2

```

alignment\_scores:

Quality: 127.00 Length: 410  
Ratio: 0.760 Gaps: 26  
Percent Similarity: 40.732 Percent Identity: 25.366

alignment\_block:  
US-09-775-693-1 x US-09-190-889A-2 ..

Align seg 1/1 to: US-09-190-889A-2 from: 1 to: 539

```

98 CCTATCTGSCCAATTTGCC.....AGAGGAGAGCTTCGAG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 ProSerHisProAlaLeuAlaLeuLeuLeuAlaLeuProAlaGlyLeuAla 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GlyProLeuAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 GGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 erProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 .....CCTACGAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 AlaArgLeuAlaProProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 TATG.....AGAGCCGCTACCTCTGCGG 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 sSerGlyHisProValAsnProValLeuProAlaAlaAlaAlaAlaAlaAla 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 CACCTCTCTGSCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....MetCysProCysGlyThrThrProPro 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 Ser.....ThrSerArgProThrSerProAla..... 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 GATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....AlaThrSerSerThrThrSerLeuAla 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 GGTCTATGTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 euGlyGlyLeuProGlyGlyCysAlaGlyCysAlaAlaAlaAlaAlaAlaAla 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 GCATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 AlaMetGlyAlaLeuGlyLeuGlySerArgGlySerThrLeuAspProGly 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 ACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 yLysProGly.....MetMetLysSerProThrAsnThr 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 CGAGGCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 hrProHisValProAlaGlyLeuProGlyAlaSerArgProProGlyLe 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 ACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 unHis.....ProGlyLeuArgAlaAlaProGlyGlyGlyPro.....P 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 CCGATGATCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 roGlyHisMetAlaAsp...ValGlyArgAsn.....GlyGlyAsp 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
708 CAACGTCACAGA...TGACACACACACACACACACACACACACACACACACAG 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GlyProLeuGlyGlyThrHisLeuArgSerGlyThrGlyGly...LeuHis 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 T.....GTACCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 scySTripLeuSerProGlyLeuArgThrAlaSerGlyAlaLeuTrpA 414

```

```

787 GAGG.....GCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 rSerGlyAlaTrpAlaAlaAlaProSerSerThr..... 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GAATGAGTCCCGAGGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 873
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 .....SerLeuArgProAlaGlyGlyProSer..... 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 CATGCTCATTTAGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 .....TrpThrGlyArg..... 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 .....GlyGlyThrLeuGlyAlaAlaGlyGlyGlyGlyGlyGlyGlyGly 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
953 .....CTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 AspLeuSerLeuGlyAlaGlyLeuGlyProCysProSerThrAlaCysTr 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
977 .....GGCCTAGCCCTGAGTGTGAT...TTGTCCG 1004
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 pHisValHisLeuLeuProGlyGlyAlaAlaAlaAlaAlaAlaAlaAlaAla 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1005 CCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 latThrLeuGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1055 CCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 lIleSerArgHisLeuGlyProProAlaAlaAlaAlaAlaAlaAlaAlaAla 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1096 CCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 ocysSerLeuSerGlyLeuGlyAlaValTrp 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq.name: /cgn2_6/prodata/2/1aa/6B_COMB.pep:US-09-190-938B-2

seq_documentation_block:
: Sequence 2, Application US/09190938B
: Patent No. 6197939
: GENERAL INFORMATION:
: APPLICANT: Farrell, Catherine L.
: Martin, Francis H.
: Yabkowski, Rachel
: TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
: FACTOR
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amgen Inc.
: STREET: 1840 De Havilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: USA
: ZIP: 91320-1789
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/190,938B
: FILING DATE: 12-No. 6197939-1998
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Mazza, Richard J.
: REGISTRATION NUMBER: 27,657
: REFERENCE/DOCKET NUMBER: A-414
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 805.447.4112
: TELEFAX: 805.499.6751
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:

```

[illegible]

```

369  roGluHisMetAlaasp...ValGlnAArgAsn.....GluGlnAasp 381
708  CAACGTCGACAGA...TGGCACCACCCACCACGCTCTTGAGGCTTCA 754
382  GluProLeuGlyGlnThrHisLeuArgSerGlyThrGlnGly...Leu11 397
755  T.....GTACGTGACGAGATCGGGCGACACGACG 786
397  scGlyTrLeuSerProGlyLeuArgArgThrHisSerGlyAlaLeuTrpA 414
787  GTGG.....GCCGATTTGACATCGTGAGAACCGCTTCATGG 823
414  rGSerGlyAlaTrpAlaAlaIaIaProSerSerTrpSerTr..... 427
824  GAATGAAGTCCGAGGTATCTACGAGACCCGACGACCATCTTAC 873
428  .....SerLeuArgProAlaGlyGlyProSer..... 436
874  CATGCTCATTTAGACATCGAGGCTTCACCATGACCGGAGACTGGCAA 923
437  .....TTTTTTTTT: 440
924  AATCAACAGCGCTGGCGTTGAATTG..... 952
441  .....GlyGlnTrpLeuGlnValLeuGlyGlyGlyHisSerGlyLeu 954
953  ...CTGACCTGGTGTATACGGGTTTAC..... 976
455  AspLeuSerLeuGlyValGlyLeuTrpProGlyProSerTrpAlaCysTr 971
977  .....GGCCTACCGCTGAGTGTGAT...TTGTCGG 1004
471  pHisValHisLeuLeuProGlyGlnAspGlyAlaLeuHisMetLeuAla 488
1005  CCACTGATCGCCCAAGTCCAGGAGAGTGTGAGAGGAAGTGCAGGCTGT 1054
488  LathrLeuGlnProGlyProProGlyGlnAlaGlyGlnGlyAlaAla 504
1055  CCGTCTCAAGGCGCAGGTGTACATCC.....TGGCCGCGGAGATCC 1095
505  TLeuSerArgHisLeuGlyProProAspProLeuProSerProProThr 521
1096  CCACTGCTCTCTACATGAGAGGCTGG 1123
521  ocGlySerLeuSerGlyLeuGlnValTrp 530

seq_name: /cgn2_6/p/tdat2/1ae/p/CTUS.COMB.pep:PCT-US95-09261-2

seq_documentation_block:
/ Sequence 2, Application PC/TUS9509261
/ GENERAL INFORMATION:
/ APPLICANT:
/ APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM
/ APPLICANT: STREET: 201 West 7th Street
/ APPLICANT: CITY: Austin
/ APPLICANT: STATE: Texas
/ APPLICANT: COUNTRY: United States of America
/ APPLICANT: POSTAL CODE: 78701
/ APPLICANT: TELEPHONE NO: (512)499-4462
/ APPLICANT: TELEFAX: (512)499-4553
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF
/ TITLE OF INVENTION: A BONE AND PROSTATE DERIVED GROWTH FACTOR
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: UNITED STATES OF AMERICA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible

```



[illegible]

```

? TITLE OF INVENTION: Protein Polymers
? NUMBER OF SEQUENCES: 135
? CORRESPONDENCE ADDRESS:
? ADDRESS: FLEHR, HOEBACH, TEST, ALBERTYTON & HERBERT
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/642,255
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: ROWLAND, Berttram I.
? REGISTRATION NUMBER: 20,015
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 494-8700
? TELEFAX: (415) 494-8771
? INFORMATION FOR SEQ ID NO: 73:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 633 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-642-255-73

alignment_scores:
  Quality: 126.50      Length: 387
  Ratio: 0.904        Gaps: 25
  Percent Similarity: 36.176  Percent Identity: 26.357

alignment_block:
  US-09-775-693-1 x US-08-642-255-73  ..

Align seg 1/1 to: US-08-642-255-73 from: 1 to: 633

105 GCCCAACATGTCGCGGAGGAGTCTGAGGAGCCAGGAGAGGAGGCAC 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
160 GAlaAlaHisGlyProAlaGlyPro..LysGlyAlaHisGlyProAla.. 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 TGAAGCTTGGGGCCAAAAGGTGTCATTGAGGATGTCAAGCAGGAGTTT 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 ..GlyProLysGlyAlaHis.....GlyProAla 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 GTGGA.....GGAGTTCATCTGCGCGGCATCCAGTCCAGCGCACTGA 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 laGlyProLysGlyAlaHisGlyProAlaGlyPro.....LysGlyAla 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 TGAAGACCGCTACCTCTGGGCACTCTCTGCAAGCCCTGCATCGGCC 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 GlnGlyProAlaGlyProGly.....GlySer 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 GCAAACAGTGGAAATCGC.....C 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 ..GlySerAlaGlyAspProGlyProGlyAlaGlnGlyP 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 CAGCGGAGGAGGCGCAAGTATGTCTCCACGCGCCACAGAAAGGGAA 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 roAlaGlyProGly.....GlySer 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 CGATCAGTCGCGTTTGAAGTCAGCTGCTACTACGCGCCCGCCAGATAA 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 ArgGlyAspPro.....GlyProGly.. 232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

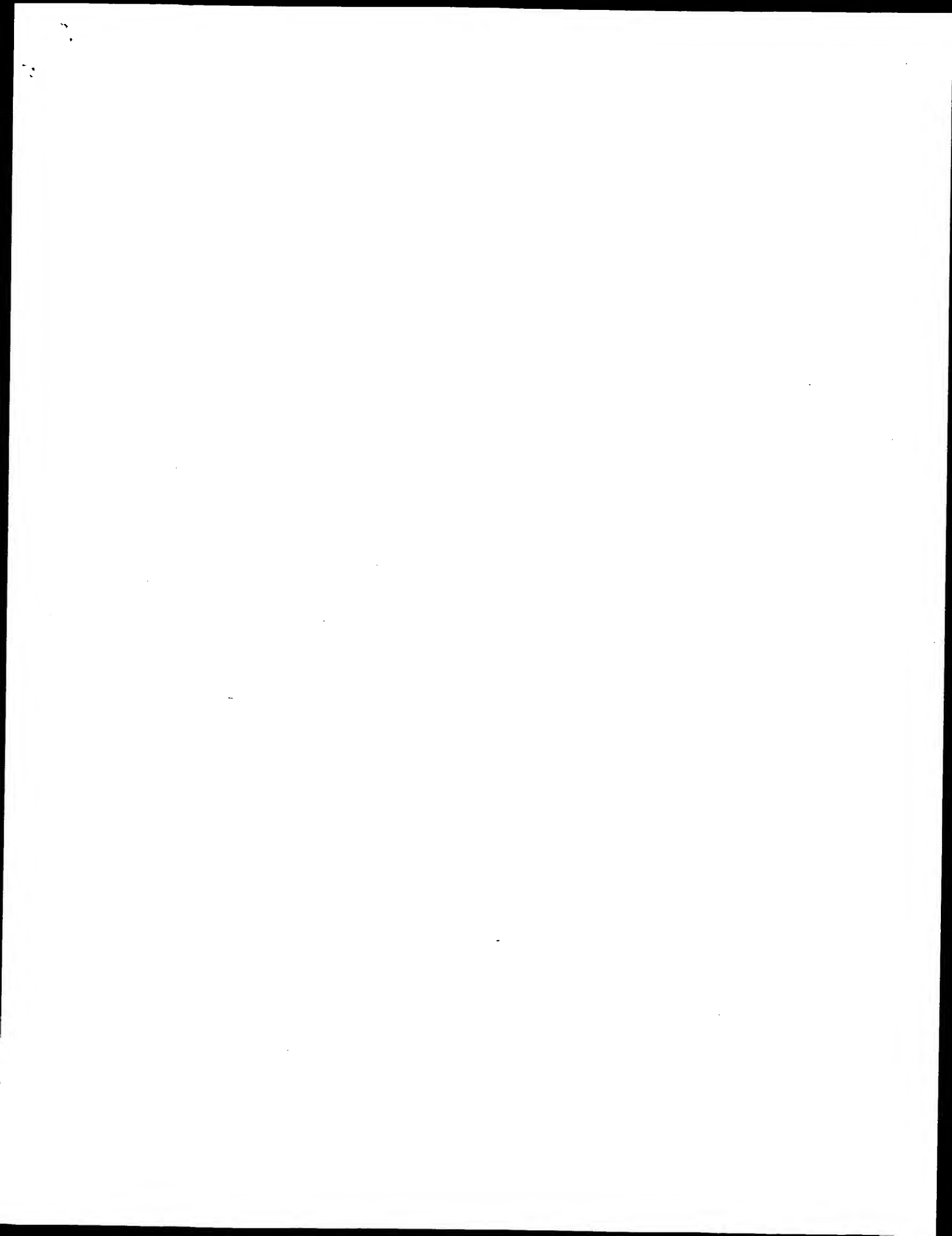
419 AGGTATGCTCTCCCTGAGAGATGCTGAATCTTACACCGGTTCAAGGC 468
233 .....AlaGlnGlyP 236
469 CGCAATGACCTGATGAGTACGCAAGCAGGAGATCCGATCCCGGT 518
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 roAlaGlyProGlySerAlaGly.....GlyAspProGlyProPro 249
519 CAGTCCCAAGAACCCCTGAGCAGTGAATGAGAACCTATGACATGAGCT 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 GAlaGlnGlyProAlaGlyProGly.....GlySer 260
569 AGGAGGCTGGAAATCTGGAGACCC.....CAAGACCAAGCGCTCA 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 raGlnGly..AspProGlyProGlyAlaGlnGlyProAlaGlyProG 276
613 GGTCTTACAGAGACCCAGGAGCC.....AGCCAAAGCCCAACAC 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 LysGlySerAlaGlyAspProGlyProGlyAlaGlnGlyProAlaGly 292
657 CCCGACATCTCGAGATCGATGTCAAAAGGGGTCCTGAGAGTGA 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
293 ProGlySerAlaGlyAspProGlyProGlyAlaHisGlyProAl 309
707 CCAAGCTCAAGATGGACACCCAGCAGACCTCTGAGCT..... 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
309 aGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly 326
750 .....CTCATGTAAGCTGAAGCAAGTGGGG...CAA 779
326 roAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaGln 342
780 GCATGGCGCTGGCGCTGATGACATGTCGAGAGACCGCTTCAATGA 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 GlyProAlaGlyProGlyGlySerAlaGlyAspPro..... 354
830 AGTCCGAGATATCTACGAGACCC.....AGCAGCGACC 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ..GlyProGlyAlaGlnGlyProAlaGlyProGlyGlySerAlaGly 371
865 ATGCTTACATGCTCATTTAGACATGAGCGCTTCACTGAGCGGGA 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 sPro.....GlyProPro 375
915 AGTGGCAAAATCAACAAGCGCTGGGCTTGAATTTGCTGAGCTGTGT 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 GAlaGlnGlyProAlaGlyProGlyGlySer..... 386
965 ATACCGGTTACGGCGCTGAGCTGATGTAATTTGCGCCAGTGCATC 1014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 .....A 387
1015 GCCAAGTCCAGAGAGAGTGAAGGAAAGTGAAGTGTGCTGCTCA 1064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
387 rGlyAspProGlyProGlyAlaGlnGly.....ProAla 399
1065 GGGGCA..GGTGTACATCTGGCGCGGAGTCCCACTGTCTCTTCA 1111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 GlyProGlyGlySerAlaGlyAspProGlyProPro.....GlyAlaGln 414
1112 ATGAGAGCTGGTGAAGATGAAGCTGAGGATTTATGACCAAGCAT 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
414 nGlyProAlaGlyProGly..GlySerAlaGlyAspProGlyProGly 430
1162 GCCACCGG 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 AlaGlnGly 433

```

Wed Feb 13 07:35:59 2002

us-09-775-693-1.rai

---





```

151 GCACGTGAAGCTGGGGCCAAAAGGCTTCATTCAGAGATGACGAGGA 200
51 AlaLeuLysLeuGlyAlaLysValPheIleGluAspValSerArgI 67
201 GTTGTGGAGAGATTTCATCGCGGCATCCGATCCGACGACGACGTATG 250
67 uPheValGluGlnPheIleTyrProAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGACACTCTTGGCAGGCGCTGCATGCGCCG 300
84 LuAspArgTyrLeuLeuGlnTyrSerLeuAlaArgProCysIleAlaArg 100
301 AACACAGTGAAGTAATCGCCAGCGGAGGCGCCAGTATGTCACAGG 350
101 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351 CGGCACAGAAAGGGAGACATCAGTCCGGTTGAGCTCAGCTGCTCT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuSerCysTyrS 134
401 CACTGGCCCCCAGATTAAGTCAATTCCTCCGAGGATGCTGTAATTC 450
134 eLeuAlaArgProGlnIleLysValIleAlaProTyrPargMetProGlnPhe 150
451 TCAACCGGTTCAAGGGCCGACATGACCTGATGGAGTACGCAAGAC 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGlnH 167
501 CGGGATTCGCATCCCGTCATCCCAAGAACCCCTGGAGCATGATGAGA 550
167 SGIlyIleProIleProValThrProLysAsnProTyrSerMetAspGln 184
551 ACCTCATGACATCAGTACGAGCTGAGATCCGAGGAGAACCCCAAGAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlnIleLeuGlnLysnProLysasn 200
601 CAAGCGCTCCAGTCTCTACAGAACACCCAGACCCAGCCAAACCCG 650
201 GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaArg 217
651 CAACACCCCTGACATTCAGATGAGATTCGAAAAAGGGTCCCTGTGA 700
217 oAsnThrProAspIleLeuGlnIleGlnPheLysGlyValProValL 234
701 AGGTGACCAACGTCAGAGATGCGACCCAGACCTCCTGGAGCTC 750
234 yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGlnLeu 250
751 TTCATGTACCTGAACGATGCGGGCAAGCATGGCTGGGCGCTATTGA 800
251 PheMetTyrLeuAsnGlnValAlaGlnLysHisGlyValGlnArgIleAs 267
801 CATCGTGGGAACCGCTTCATTCGATGAAGTCCGAGGTATCTACGAGA 850
267 rIleValGlnAsnArgPheIleGlyMetLysSerArgIlyIleTyrGlnT 284
851 CCCGAGGAGGACCATCCCTTAACATGCTCATTTAGACATGAGCGCTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 300
901 ACCATGACCGGGAAGTGGCCAAATCAACAGGCTGGGCTTGAATTT 950
301 ThrMetAspArgGlnValArgLysIleLysGlnGlnIleuGlnLeuLysPh 317
951 TGCTGAGGTGGTATACCGGTTTACGGCTTACGCTGAGTGTGAATTTG 1000
317 eAlaGlnLeuValTyrThrIleLeuArgProSerProGlnCysGlnPheV 334
1001 TCAGCACTGATGCGCAAGTCCAGAGGAGTGAAGGAAAGTGCAG 1050
334 aArgHisCysIleAlaLysSerGlnGlnTyrValGlnLysValGln 350
1051 GTGTCCTCTCAAGGGCCAGGTGTACATCTCGCGGAGAGTCCCACT 1100

```

```

|||||
351 ValSerValLeuLysGlnValTyrIleLeuGlyArgGlnSerProLe 367
1101 GTCTCTTACATAGAGAGCTGGTGAACATGACGTCAGGCTGATTTATG 1150
367 uSerLeuTyrAsnGlnGlnLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCAACTATGCGCACCGGTTTCATCAATCAATCAATTCCTCAGCTGAAG 1200
384 LuProThrAspAlaThrGlnPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATTCATCGTCTCCAGAGCAAGGTCACTGCGCAAA 1236
401 GlnTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: p1r1:AUFRS
seq_documentation_block:
argininosuccinate synthase (EC 6.3.4.5) - rat
N:Alternate names: citrulline--aspartate ligase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
C:Accession: S01440
R:Sunh, L.C.; Morris, S.M.; O'Brien, W.E.; Beaudet, A.L.
Nucleic Acids Res. 16, 9352, 1988
A:Title: Nucleotide sequence of the cDNA encoding the rat argininosuccinate synthetas
A:Reference number: S01440; M01D:89016648
A:Accession: S01440
A:Molecule type: mRNA
A:Residues: 1-412 <SUR>
A:Cross-references: EMBL:X12459; NID:955766; PIDN:CA03099.1; PID:955767
C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline a
C:Superfamily: argininosuccinate synthase
C:Keywords: arginine biosynthesis; homotetramer; ligase; urea cycle
F:149,153/Binding site: Mg-ATP (Glu, Arg) #status predicted
alignment_scores:
Quality: 2086.00 Length: 412
Ratio: 5.113 Gaps: 0
Percent Similarity: 99.029 Percent Identity: 96.845
alignment_block:
US-09-775-693-1 x AUFRS
Align seg 1/1 to: AUFRS from: 1 to: 412
1 ATGTCCAGCAAGGCTCCGTTGCTTGGCTTACAGTGGCGGCTGACAC 50
|||||
1 MetSerSerLysGlnSerValValLeuAlaTyrSerGlyGlyLeuAspThr 17
51 CTCGTGCAATCCGCTGGTGGCTGAAGGAACAGGCTATGAGCTCATTCGCT 100
|||||
17 rSerCysIleLeuValThrPheLysGlnGlnGlyTyrAspValIleAlaThr 34
101 ATCTGGCCACATTTGGCCAGAGAAAGACTTCGAGAGACGAGAGAGAG 150
|||||
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaArgLysLys 50
151 GCACGTGAAGCTTGGGGCCAAAAGGCTTCATTCGAGAGTCTGACAGAGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysValPheIleGlnAspValSerLysGln 67
201 GTTGTGGAGAGATTTCATTCGCGGCATCCGATCCGACGACGCTGATG 250
|||||
67 uPheValGlnGlnPheIleTyrProAlaValGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGACACTCTTGGCAGGCGCTGCATGCGCCG 300
|||||
84 LuAspArgTyrLeuLeuGlnTyrSerLeuAlaArgProCysIleAlaArg 100
301 AACACAGTGAAGTAATCGCCAGCGGAGGCGCCAGTATGTCACAGG 350
|||||

```

```

101 LysGlnValGluIleAlaGlnArgGluGlyAlaValSerHisG1 117
351 GGGCCAGGAAGGGGAGACGATCGAGTCGGTTGAGCTGAGCTGTAAT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuThrCysTyrS 134
401 CAGTGGCCCCCAGATAAAGTCATGCTGCTCCCTGGAGAGATGCTGAAATC 450
134 erLeuAlaProGlnIleLeuValIleAlaProTrpArgMetProGluPhe 150
451 TACAAACGGTTCAAGGGCCGCAATGAGTGGATGAGTCCGCAAGCAACA 500
151 TyrAsnArgPheLysGlyArgAsnSpmLeuMetGluTyrAlaLysGlnH1 167
501 CGGATTCCTCCATCCCGGTCACTCCCAAGACCCGTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysSerProTrpSerMetAspGluA 184
551 ACCTCATGACATCAGCTACGAGGCTGGAATCCCTGGAGACCCCAAGAC 600
184 snLeuMetHisLeuSerTyrGluAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAGGCCCTCCAGGCTCTACACGAAGCCAGACCCAGCCCAAGCCCC 650
201 GlnAlaProProGlyLeuTyrThrIlysthrglnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATTCCTGAGATGAGTTCAAAAAGGGCTCCCTGTA 700
217 oAsnThrProAspValLeuGlnIleGluPheLysLysGlyValProVal 234
701 AGGTGACCAAGCTCAAGAGTGGACACCCACCACTCTGAGAGCTC 750
234 yValThrAsnValLysAspGlyThrHisSerThrLeuAsnSpmLeu 250
751 TTCATGTACTGAACGAAGTGGGGGCAAGCATGCGTGGCCGATTTGA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValAlaArgLys 267
801 CATGTGGGAACCGCTTCATGGATGAAGTCCGAGGATCTACGAGA 850
267 rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGlnT 284
851 CCCAGAGACACCATCTTACCATGCTCTATTAGACATGAGGCTTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300
901 AACATGACCGGGAAGTCCGCAAAATGAACAAGCCTGGGCTTGAATTT 950
301 ThrMetAspArgGlnValArgLysIleLysGlnGlyLeuGlyLeuLysPh 317
951 TGGTGAAGTGGTATACCGGTTACGGCTAGCCCTGAGCTGAAATTTG 1000
317 eAlaGlnLeuValTyrThrGlyPheTrpHisSerProGluCysGluPheV 334
1001 TCCGCGCACTGATCGCAAGTCCAGAGAGCATGAGGAAGGAAGTCCAG 1050
334 aArgHisCysIleAspLysSerGlnIleArgValGlnGlyLysValGln 350
1051 GTGTCCGCTCCAGGGCCGAGTGTACATCTCCGCGGAGAGTCCCACT 1100
351 ValSerValPheLysGlyValTyrIleLeuValArgGlnSerProLe 367
1101 GTCTCTACATGAGAGCTGTGTAGCATGACATGAGGCTGATATG 1150
367 uSerLeuTyrIleAsnGlnGlnLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCAACTGATGCCACGGGTTCATCAACATCAATTCCTCAGCTGAG 1200
384 lUpProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATCTCTCCAGAGCAAGGTCACTGCCA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412

```

```

seq_name: p1r1:AJMSRS
seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) - mouse
  N:Alternate names: citrulline--aspartate ligase
  C:Species: Mus musculus (house mouse)
  C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
  C:Accession: J00463
  R:Stuh, L.C., Beaudet, A.L., O'Brien, W.E.
  Gene 99, 181-189, 1991
  A:Title: Molecular characterization of the murine argininosuccinate synthetase locus.
  A:Reference number: J00463, M01D:91216457
  A:Accession: J00463
  A:Molecule type: mRNA
  A:Residues: 1-412 <SDR>
  A:Cross-references: GB:M1690; NID:g192068; PIDN:AAA37266.1; PID:g309111
  C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline a
  C:Genetics:
  A:Gene: ASS
  A:Introns: 35/3; 58/3; 121/3; 142/3; 495/3; 189/2; 199/3; 230/1; 258/2; 280/1; 324/1;
  C:Superfamily: argininosuccinate synthase
  C:Keywords: arginine biosynthesis; homotetramer; ligase; urea cycle
  F:149/133/Binding site: Mg-ATP (Glu, Arg) #status predicted

```

```

alignment_scores:
  Quality: 2073.00      Length: 412
  Ratio: 5.093          Gaps: 0
  Percent Similarity: 98.786  Percent Identity: 96.117

```

```

alignment_block:
  us-09-775-693-1 x AJMSRS

```

```

Align seg 1/1 to: AJMSRS from: 1 to: 412

```

```

1 ATGTCCAGGAAGGCTCCGCTGCTGCTGCTACAGTGGCGGCTGAGAC 50
1 MetSerLysGlySerValValLeuAlaTyrSerGlyLysLeuAspTh 17
51 CTCGTGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100
17 rSerCysIleLeuValThrPheLysGlnGlnGlyTyrAspValIleAlaT 34
101 ATCTGGCAACATGTGGCCAGAGAGACTTGGAGAGACCCAGAGAGAG 150
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaArgLysLys 50
151 GCAGTGAAGCTTGGGGCCAAAGGCTGCTTACATGAGATGTCAGCGCA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerLysG1 67
201 GTTGTGGAGAGATTCATCTGCGCGGCGGCTGCTGCTGCTGCTGCTG 250
67 uPheValGlnGluPheIleThrProAlaValGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTGGGCACTCTCTTGGCAGGCGCTGCATGCGCGC 300
84 lAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100
301 AAACAAGTGAATGCGCCAGCGGAGGAGGCGCAAGTATGTCGCCAGG 350
101 ArgGlnValGluIleAlaGlnArgGlnValAlaLysTyrValSerHisG1 117
351 GCGCACAGGAAGGGGAGAGATCAGGTCGGTTGAGCTGAGCTGTAAT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuThrCysTyrS 134
401 CAGTGGCCCCCAGATAAAGTCATGCTGCTCCCTGGAGAGATGCTGAAATC 450
134 erLeuAlaProGlnIleLysValIleAlaProTrpArgMetProGluPhe 150
451 TACAAACGGTTCAAGGGCCGCAATGAGTGGATGAGTACGCAAGCAACA 500

```

```

151  TYSNAPRPHLYSGIYARGASNPSEUWETGLUTYRALALYSINH 167
501  CGGGATTCCTCCATCCGGTCACTCCCAAGAACCGGTGAGATGATGACA 550
167  SGIYILEPROVALTHPROLYSSETPROTPSERMETASPGIUA 184
551  ACCATGACATACAGTACAGAGCTGGAATCCGTGAGAACCCCAAGAAC 600
184  SNEUWETHISLESERTYRGUAGIYILEUENLUNSPROLYSASN 200
601  CAAGCGCTCCAGTCTCTACAGAGACCCAGAACCCCAAGAACCC 650
201  GINALEPROVALTHPROLYSSTHGLINASPPOVALALYSALAPR 217
651  CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGTCCCTGGA 700
217  OASNSETPROASPVALLLEUGLULLEGLUPHELYSYSGIYVALPROVAL 234
701  AGGTGACCAACCTCAAGATGCGACACCCAGACCTCTTGAGGCTC 750
234  YSVALTHRSNILEYSPGLYTHRTYRGTHTHRSERLEUGLULEU 250
751  TTCATCTACCTGACGAAGTCCGGGCAAGCATGGCGGCGCTGATGGA 800
251  PHMETYRLEUASNGLUVALAAGLYSHISGLIYVALIGIYARGILEAS 267
801  CATCGGGAGAACCGCTTCATTTGAATGAAGTCCGAGGTATCTACAGA 850
267  PILEVALIGUASNARGPHEILEGLYMETLYSSERTARGIYLEYRGUT 284
851  CCCGACGACGACCATCTTACCATGCTGCTCAATTAGACATCGAGCCCTC 900
284  HPROVALGLYTHRTLEUUTYRHSALAHISLEUASPILEGUVALAPHE 300
901  ACCATGACCGGGAAGTCCGCAAAATCAACAGAGCTGGGCTTGAAT 950
301  THRMETASPARGLUVALARGLYSILEYSGINLILEUGLILEUENLUSPH 317
951  TCGTAGTGATACCGGTTACGAGCTAGCCCTGAGTGTGAAATTTG 1000
317  EALAGLILEUVALYTHRTYRGTHTHRTYRHSERPROGLICYSGIUPHEV 334
1001  TCCGCGCATGTCATGCCAAGTCCAGAGAGGAGTGAAGGGAAGTGCAG 1050
334  ALARGHISCYSEILEGLINLYSSERTINLUARGYVALIGIUGLYSVALGIN 350
1051  GTGTCGCTCAAGGCGCAAGGTGTACATCTCGCGGAGGAGTCCCGCAT 1100
351  VALSERVALPHLYSGIYGLINVALYTHRTLEUGLYARGGLUSERPROLE 367
1101  GTCTCTACATGAGAGCTGGTGGAGCATGAGCGAGGAGTATATG 1150
367  USERLEUTYRASNGLUGLILEUVALSERMETASNVALINGIYASPTYG 384
1151  AGCAATGATGCCACCGGGTTCATCAACATCAATTCCTCAGGCGGAG 1200
384  LUPROLLIASPALATHRTGLYPHEILEASNILEASNSERLEUARGLEUYS 400
1201  GAATATCATGCTCTCCAGAGCAAGTCACTGCCAA 1236
401  GLUTYRHSARGLEUENLILEYSGIYVALTHRTALALYS 412

```

seq\_name: p1r1:AJBORS

seq\_documentation\_block:

argininosuccinate synthase (EC 6.3.4.5) - bovine

N:Alternate names: citrulline--aspartate ligase

C:Species: Bos primigenius taurus (cattle)

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 23-Feb-1997

C:Accession: A33986

R:Demis, J.A.; Healy, P.J.; Beaudet, A.L.; O'Brien, W.E.

Proc. Natl. Acad. Sci. U.S.A. 86, 7947-7951, 1989

A:Title: Molecular definition of bovine argininosuccinate synthase deficiency.  
A:Reference number: A33986; MUID:90046714  
A:Accession: A33986  
A:Molecule type: mRNA  
A:Residues: 1-412 <DPN>  
A:Cross-references: GB:M26198  
C:Comment: This enzyme catalyzes the formation of argininosuccinate from citrulline and aspartate.  
C:Superfamily: argininosuccinate synthase  
C:Keywords: arginine biosynthesis; homotetramer; ligase; urea cycle  
F:149,153/Binding site: Mg-ATP (Glu, Arg) #status predicted

alignment\_scores:  
Quality: 2061.00 Length: 412  
Ratio: 5.039 Gaps: 0  
Percent Similarity: 99.272 Percent Identity: 94.903

alignment\_block:  
US-09-775-693-1 x AJBORS ..

Align seg 1/1 to: AJBORS from: 1 to: 412

```

1  ATGTCAGCAAGAGCTCCGTGTTCTGAGCTACAGTGGCGGCTGACAC 50
1  MetSerGlyLysGlySerValValLeuAlaTyrSerGlyLeuAspThr 17
51  CTCTGATCTCTGTTGGCTGAGAGAACAGGCTATGACGTCATTCCT 100
17  rSerGlyLeuValAlaTrpLeuLysGluGlnGlyTyrAspValIleAla 34
101  ATCTGGCCCAACATTGGCCCAAGAGAACCTTCAGAGACCCAGAGAAC 150
34  YrLeuAlaAsnIleGlyGlnGlnLysPheGlnGlnGlnAlaArgLys 50
151  GCATGAAAGCTTGGGCGCAAAAGTGTTCATGAGAGATGACGAGGGA 200
51  AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspIleSerLys 67
201  GTTGTGAGAGATTCATCTGGCGGCGCATCAGTCCAGGCACTGATG 250
67  UpheValGlnGlnPheIleTrpProAlaIleGlnSerSerAlaLeuTyr 84
251  AGGACCCCTACCTCTGGGACCTCTGTCAGGCGGCTGACGCGCCG 300
84  LuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100
301  AAACAACTGGAATGCCCAAGCGGAGGCGCCCAAGTATGTGCCACGG 350
101  LysGlnValIleIleAlaGlnArgGlnGlyAlaLysTyrValSerHis 117
351  CGGCACGAGAAAGGGGACGATGAGTCCGGTTTGAGCTCAGCTGTACT 400
117  YAlaThrGlyLysGlyAsnAspGlnIleArgPheGluLeuThrCysTyr 134
401  CACTGGCCCCCAGATPAAAGTCAATTCCTCGAGAGTGGCTCAATTC 450
134  exLeuAlaProGlnIleLysValIleAlaProTrpArgMetProGluHe 150
451  TACAAACGGTTCAGGGCGGCAATGACCTGATGAGAGTGCAGAAAGCA 500
151  TyrAsnArgPheGlnIleYrAsnAspLeuWetGluTyrAlaAsnGlnH 167
501  CGGGATTCCTCCATCCGGTCACTCCCAAGAACCCGGAGAGATGAGA 550
167  SGIYILEPROVALPROVALTHPROLYSSETPROTPSERMETASPGIUA 184
551  ACCATGACATACAGTACAGAGCTGGAATCCGTGAGAACCCCAAGAAC 600
184  SNEUWETHISLESERTYRGUAGIYILEUENLUNSPROLYSASN 200
601  CAAGCGCTCCAGTCTCTACAGAGACCCAGAACCCCAAGAACCC 650
201  GINALEPROVALTHPROLYSSTHGLINASPPOVALALYSALAPR 217

```





810 GAACGGCTTCATTGGAATGAGTCCGAGGATATCTACGAGACCCAGAC 859  
 268 uaaanargpheiieglyileysserarglyvaltyrghlthrproglya 285  
 860 GCACCATCTTTACATGCTCATTTAGACATCGAGCCCTTCACATGAC 909  
 285 lathrilleutrpillealhisargaspheuglucglyilethrmetsp 301  
 910 CGGAGTGGCGCAAAATCAACAGAGCCCTGGCTGAAATTTGCTGAGCT 959  
 302 lysgluvalmethisluenarvaspmetleualaprolysphealaglu 318  
 960 GGTGTATACCGGTTTACGCGCTTACGCTGAGTGTGATTTGTCGCCACT 1009  
 318 uiletyrasnglyphetrpheserproglumetclupheleuualaa 335  
 1010 GCATGCGCCAGTCCAGAGGAGCATGTGAAGGAAAGTGCAGGTGCGTC 1059  
 335 laphnrglysalaglnlunasnvalthrlylvalthrvalserile 351  
 1060 CTCAAGGGCCAGGTGTACATCTCGCGCGGAGGCCACCTGCTCTCTA 1109  
 352 tyrllysglyasnvalmetprovalalargtyrserprotyrserleuty 368  
 1110 CAATAGAGAGCTGGTGAAGCATGAACGTGCAGAGGTGATTATAGCCAACTG 1159  
 368 rasnprogluenseisermetaspvalgluglylphesapalathra 385  
 1160 ATGCCACCGGGTTTCATCAACATCAATTCCTCGAGCGGAGGCAATATCAT 1209  
 385 spserlysglyphelileasnlelhisalaleuargleuys...ValHis 400  
 1210 CCTCTCCAGAGCAAG 1224  
 401 Glnleuvallylys 405

seq\_name: p1r2:T40457  
 seq\_documentation\_block:  
 arglino succinate synthase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T40457  
 R:Lyne, M.; Rajandream, M.A.; Barrall, B.G.; Baker, S.; Mungall, K.  
 submitted to the EMBL Data Library, November 1998  
 A:Reference number: 221931  
 A:Accession: T40457  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-410 <LYN>  
 A:Cross-references: EMBL:AL034382; PIDN:CAA22280.1; GSPDB:GN00067; SPDB:SPBC428.05c  
 A:Experimental source: strain 972n-; cosmid c428  
 C:Genetics:  
 A:Gene: SPDB:SPBC428.05c  
 A:Map position: 2  
 C:Superfamily: arglino succinate synthase

alignment\_scores:  
 Quality: 1083.00 Length: 395  
 Ratio: 3.384 Gaps: 4  
 Percent similarity: 81.013 Percent identity: 54.430

alignment\_block:  
 US-09-775-693-1 x T40457

Align seg 1/1 to: T40457 from: 1 to: 410

22 GTTCGGCTTACAGTGGCGCTGAGACCTGTCATCTCTGCTGGCT 71  
 |||||||  
 9 ValneuAlarYserGlyGlyLeuAspThrSercyslleuAlarTiple 25

72 GAAGAAACAAGGCTATAGCTATTCCTGCTATGCGCAACATTTGGCCAGA 121  
 | |||||  
 25 uilegluugluyttrpbluvalillecyslyrmetalaasnvalglyling 42  
 122 AGGAGACTTCGAGAGAGCCAGAGAGGCAAGGCACTGACCTGGGGCCAA 171  
 : |||||  
 42 luclunsprrpsalaaalarglulysAlaleuLysvalGlyAlalys 58  
 172 AAGGTGTTCATGAGAGATGTACAGAGGAGTTTGTGAGAGATTCATCTG 221  
 |||||  
 59 LysValTyfValGluAspLeuArgGluGluPheleAsnAspThrVal11 75  
 222 GCGGCGCATCCAGTCCAGCGCATGTATGAGAGAGCCCTACTCTGGGCA 271  
 |||||  
 75 eproAlaIaIaGlnAlaAsnAlaIleTyfGluAsnValTyfLeuLeu1y 92  
 272 CCTCTCTTGGCAGAGCCCTGCATCGCGCCGCAACAGTGAATTCGCCAG 321  
 |||||  
 92 hrserleuAlarprolelleAlarargGlnIleGlnIleAlaGlu 108  
 322 CGGAGGGGGCCAGATATGTCTCCAGCGGCCACAGAGAGGAGCA 371  
 |||||  
 109 LysGluAsnGlyIleAlaValSerHisGlyCysThrGlyGlyGlyAsnAs 125  
 372 TCAGGTCGGTTTGAAGCTCACTGCTACTGCTGCGGCCCGCCAGATTAAG 421  
 |||||  
 125 pgluValarphelgluLeuAlaTyfAlaLeuLysProAspValGlnv 142  
 422 TCATTTGCTCCGCTGAGAGATGCTGATTTCTACAAACGGTTTCAAGGCCG 471  
 |||||  
 142 alileAlarprotrpArgLeuProvalPhehegluArgPhealaglyarg 158  
 472 AATGACCTGATGAGTACGCAAGCAACAGGAGATTCCTCCATCCGCTGAC 521  
 |||||  
 159 LysAspLeuLeuGluTyfAlaAlaAlaLysGlyIleProvalThrGln1 175  
 522 TCCCAAGAACCGGTGAGAGATGATGAGAACCTCATGACATCAGCTAG 571  
 |||||  
 175 rThrLysLysProtrpsermetaspGluAsnIleValHisCysSertyrg 192  
 572 AGGCTGGAATCTCGAGAACCCCAAGAACCAAGCGCTCCAGTGTCTAC 621  
 |||||  
 192 luAlaglyIleleuGluAspProsermetThrProProLysAspmetrtp 208  
 622 ACAGAACCCAGAGCCAGCCAAAGCCCAACACCCCTGACATCTCTGA 671  
 |||||  
 209 LysLeuThrValAspProLysAspAlarProAspGluValGluGluLeu 225  
 672 GATCGAGTTCAAAAAAGGGCTCTGTGAGGTGACCAAGCTCAAGATG 721  
 |||||  
 225 rIleHisPheglulysGlyAlarProThrLysLeu...GluCysLysaspG 241  
 722 GCACCAACCAGACCTCTTGAAGCTTCATGATCTGACCTGAAGAGATG 771  
 |||||  
 241 ly...ThrPheSerGlyValValSerIlePheTyfGlnLeuAsnAlaIle 256  
 772 GCGGCAAGCATGGCTGGCGCTATGACATCTGTGAGAACCGCTTCAT 821  
 |||||  
 257 AlaArgArGAsnGlyValGlyArgIleAspIleValIunsnrgrPhe 273  
 822 TGGAAATGAAGTCCGAGGTATCTACGAGACCCAGAGCCAGCATCTCTT 871  
 |||||  
 273 rGlyLeuLysSerArgGlyCysTyfGluThrProGlyLeuThrIleleu 290  
 872 ACCATGCTCATTTAGACATGAGGCTTCAACATGACGCGGAGATGCGC 921  
 |||||  
 290 rGThrAlaHisMetAspLeuGluGlyLeuThrMetGluArgGluValArg 306  
 922 AAAATCAAA...CAAGCCTGGGCTTGAATTTGCTGAGAGTGTATATC 968  
 |||||  
 307 AlaLeuArgaspGlnPheValThrPheAsnLeuAlaLysIleleuTyfAs 323  
 969 CGGTTACGGGCTAGCCCTGAGTGTGAATTTGTCCGCCACTGCATCGCCA 1018

```

|||||
323 nGlylnhpheserProcythrArgMetLeuAlaAlaasnAsnV 340
1019 ACTCCAGAGCGAGCGAGGAAAGCGAGGCGCGCTCAAGGCG 1068
340 aISerGlnGlnValAlaSnGlnValValLysLysSerValTyrLysGly 356
1069 CAGGCTGATCCTCGCGCGGAGTCCCGCATGCTCTCAATGA 1115
357 AsnValThrValLeuGlyArgLysSerAspThrAlaHisLeuTyrAspGly 373
1116 GGAGCTGTGAGCATGAACTGCGAGGTGATTAAGCCCACTGCGCA 1165
373 uLysLysSerSerMetAspGlnLeuGlnGlyLysAspProThrThrPrrns 390
1166 CCGGTTTCATCAACATCATTCCTCCAGCGCTGAG 1200
390 eRgLYpHeIleGlnIleGlnLysSerMetArgLeuArg 401
seq_name: p1r1:AJBYRS

```

```

seg_documentation_block:
arginosuccinate synthase (EC 6.3.4.5) - yeast (Saccharomyces cerevisiae)
N:Alternate names: Citrulline-lyase; protein O1228; protein YOL058w
C:Species: Saccharomyces cerevisiae
C>Date: 30-Jun-1992 #sequence_revision 08-Mar-1996 #text_change 23-Mar-2001
C:Accession: S59291; J00779; S61721; S66750; S41983
R:Manhaupt, G.; Vetter, I.; Schwarzlose, C.; Mitzel, S.; Feldmann, H.
submitted to the EMBL Data Library, August 1995
A:Description: Analysis of a 26kb region on the left arm of yeast chromosome XV.
A:Reference number: S59285
A:Accession: S59291
A:Molecule type: DNA
A:Residues: 1-420 <FEEL>
A:Cross-references: EMBL:X91067; NID:9984177; PIDN:CAA62528.1; PID:9984184
R:Van Vliet, F.; Crabbe, M.; Boyen, A.; Tricot, C.; Stalon, V.; Falmagne, P.; Nakamura,
Gene 95, 99-104, 1990
A:Title: Sequences of the genes encoding arginosuccinate synthetase in Escherichia coli
A:Reference number: J00778; M01D:91071613
A:Accession: J00779
A:Molecule type: DNA
A:Residues: 1-26, 'AT', '29-47', 'VL', '50-60', 'GGS', '65-168', 'F', '170-315', 'L', '317-420' <VAND>
A:Cross-references: GB:M35237; NID:9171084; PIDN:AAA4437.1; PID:9171085
R:Manhaupt, G.; Vetter, I.; Schwarzlose, C.; Mitzel, S.; Feldmann, H.
Yeast 12, 67-76, 1996
A:Title: Analysis of a 26 kb region on the left arm of yeast chromosome XV.
A:Reference number: S61715; M01D:96381248
A:Accession: S61721
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-420 <MAN>
A:Cross-references: EMBL:X91067; NID:9984177; PIDN:CAA62528.1; PID:9984184
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Feldmann, H.; Manhaupt, G.; Vetter, I.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66743
A:Accession: S66750
A:Molecule type: DNA
A:Residues: 1-420 <FEW>
A:Cross-references: EMBL:Z74800; NID:91419870; PIDN:CAA99067.1; PID:91419871; GSPDB:GN00
A:Experimental source: strain 5288C
R:Crabbe, M.; Seneca, S.; Devos, K.; Glansdorff, N.
Curr. Genet. 13, 113-124, 1988
A:Title: Arginine repression of the Saccharomyces cerevisiae ARG1 gene.
A:Reference number: S41983; M01D:88223481
A:Accession: S41983
A:Molecule type: DNA
A:Residues: 1-26, 'AT', '29-47', 'VL', '50-57' <GRA>
A:Cross-references: EMBL:X07070; NID:93371; PIDN:CAA30106.1; PID:93372
C:Genetics:
A:Gene: SGD:ARG1; MIPS:YOL058w
A:Cross-references: SGD:S0005419; MIPS:YOL058w
A:Map position: 15L

```

C:Superfamily: arginosuccinate synthase  
 C:Keywords: arginine biosynthesis; homotetramer; lyase  
 F:148,152/Binding site: Mg-ATP (Glu, Arg) #status predicted

alignment\_scores:  
 Quality: 1069.00 Length: 414  
 Ratio: 3.351 Gaps: 4  
 Percent Similarity: 77.053 Percent Identity: 50.483

alignment\_block:  
 US-09-775-693-1 x AJBYRS ..

Align seg 1/1 to: AJBYRS from: 1 to: 420

```

7 AGCAAGGCTCGGCTGTTGCTGCTACAGTGGCGGCTGACACTGCTG 56
|||||
2 SerLysGlyLysValLysLeuAlaTyrSerGlyLysLeuAspThrSerVa 18
57 CATCTGCTGTGCTGAGAGCAAGAGCTATGAGTCTCATTCCTGCTG 106
18 LLeuAlaThrLeuAlaSpGlnGlyLysValValAlaValAlaPheVa 35
107 CCAACATTCGCGCAGAGAAAGCTTCAGAGCCAGAGAGGCACTG 156
|||||
35 LAsnValGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 51
157 AAGCTTGGGCGCAAAAGGTTCATTGAGTATGAGTATGAGGAGTTCT 206
52 LysLeuLysLysLysLysPheValLysValAspCysArgGlnAspPheVa 68
207 GGAGAGTTCATCTGCGCGCGCCATCCAGTCCAGCGCTGATGAGAGC 256
68 LlysAspLysLeuPheProAlaValGlnValAlaValAlaValTyrGlnAspV 85
257 GCTACCTCTGCGGCGCTCTCTGCGAGGCTGATCGCGCGCAACAA 306
85 aTyrLeuLeuGlyThrSerLeuAlaArgProValAlaValAlaValGln 101
307 GTGGAATTCGCCAGCGGAGGAGGCGCAATATGCTGCTCCAGCGCC 356
102 LLeuValAlaLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 118
357 AGGAAGGGAGAGCATGCTGCGGTTGAGCTGAGCTGACTGCTG 406
118 rGlyLysGlnLysAspGlnLysLysLysLysLysLysLysLysLys 135
407 CCCCCAGATAAAGTCTATTCCTGCGAGAGATGCTGATTTCTACAAC 456
135 yProAspValLysCysLysThrProThrPargMetProGlnPhePheGlu 151
457 CGGTTCAAGGCGCGCAATGACCTGATGAGTACGCAAGCAAGCGGAT 506
152 ArgPheAlaGlyArgLysAspLeuLysPyrAlaAlaGlnLysGlyL 168
118 rGlyLysGlnLysAspGlnLysLysLysLysLysLysLysLysLys 135
407 CCCCCAGATAAAGTCTATTCCTGCGAGAGATGCTGATTTCTACAAC 456
135 yProAspValLysCysLysThrProThrPargMetProGlnPhePheGlu 151
507 TCCCATCCCGGCTCACTCCCAAGAACCCGTGAGCATGATGATGAT 556
168 eProValAlaGlnThrLysAlaLysProThrPargMetProGlnPheGln 185
557 TGCACATCAGCTACGAGGCTGATGATGATGATGATGATGATGATG 606
185 LHisLysSerLysGlnAlaGlyLysLysLysLysLysLysLysLys 201
607 CCTTCAGGCTCTACAGAGAACCCAGAGCCAGGCAAGCCCGCCAC 656
202 ProLysAspMetThrLysLeuLysValAspPrometAspAlaProAspG 218
657 CCCTGACATTCGAGATGAGTTCAAAAGGCGCTGCTGAGGCTGAG 706
218 nProGlnAspLeuThrLysPheGlnLysGlnLysLysLysLysLys 235
707 CCAAGTCAGAGATGGACCCAGCCAGCAAGCTCC.....TTG 744
|||

```

235 hTyrThrAspAsnLysThrSerLysGluValSerValThrLysProLeu 251  
 745 GAGCTCTCATCTACCTGACCAAGATCGCGCAAGCATGGCTGGCGG 794  
 252 AspValPheLeuAlaIAserAsnLeuAlaArgAlaAsnGlyValGlyAr 268  
 795 TATTGACATCGTGAGAACCGCTTCATTGTAATGAATCCGAGGATCT 844  
 268 gileAspIleValGluAspArgTyrIleAsnLeuLysSerArgGlyCysT 285  
 845 ACAGACCCCGACAGACCATCTTACCATGCTCATTTAGACATCGAG 894  
 285 yrlGluInlAlaProLeuThrValIleArgLysAlaHisValAspLeu 301  
 895 GCGTTCACCATGACCGGGAAGTGGCAAAATCAACAAAGGC...CTGGG 941  
 302 GlyLeuThrLeuAspLysGluValArgLLeuLysAspSerPheValTh 318  
 942 CTTGAAATTTGCTGAGCTGTGTATACCGGTTTACGGCTAGCCCTAGT 991  
 318 rProAsnTyrSerArgLeuIleTyrAsnGlyPheLeuLeuHisProGlu 335  
 992 GTGAATTTGTCGCCACATGACATCCCAAGATCCGAGAGCGAGTGAAGC 1041  
 335 ysgLutTyrIleArgSerMetIleGlnProSerGlnAsnSerValAsnGly 351  
 1042 AAAGTCAGAGTGCTCCCTCTCAAGAGCGAGTGTACATCTCGAGCGGA 1091  
 352 ThrValArgValArgLeuTyrGlyValAsnValIleIleLeuLysrge 368  
 1092 GTCCCACTG...TCTCTACAAATGAGAGCTGCGTACGATGACGTGC 1138  
 368 rThrLysThrGluLysLeuTyrAspProThrGluSerSerMetAspGlu 385  
 1139 AGGTGATTTATGACCAATGATGATCCAGCGGTTTCATCAACATCAATCC 1188  
 385 eutHrGlyPheLeuProThrAspThrThrGlyPheIleAlaIleGlnAla 401  
 1189 CTCAGCGTGAAGAAATATCATCTGCTCCAGACAGCAAGTCACT 1230  
 402 IleArgIleLysLysTyrGly.....GluSerLysLysThr 413

seq\_name: p1r2:D75490

seq\_documentation\_block:  
 argininosuccinate synthase - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Smith, H.O.; Vamathevan, J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Ma  
 Science 286, 1571-1577, 1999  
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: D75490  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-402 <WHI>  
 A:Cross-references: GB:AE001924; GB:AE000513; NID:g6458372; PIDN:AAF10250.1; PID:g645837  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0674  
 A:Map position: 1  
 C:Superfamily: argininosuccinate synthase

alignment\_scores:  
 Quality: 1034.00 Length: 411  
 Ratio: 3.272 Gaps: 6  
 Percent Similarity: 76.886 Percent Identity: 50.122  
 alignment\_block:

US-09-775-693-1 x D75490 ..  
 Align seg 1/1 to: D75490 from: 1 to: 402

7 AGCAAGAGCTCCGAGTGTCTGCTACAGTGGCGGCTGACACCTGCTG 56  
 2 SerLysGluLysIleValLeuAlaTyrSerGlyGlyLeuAspThrSerI 18  
 57 CATCCCTCGTGTGCTGAAG...GAACAAGCTATGAGCTGATTCCTATC 103  
 18 eileLeuLysTyrLeuGlnThrGlnArgAsnTyrAspValValCysPhe 35  
 104 TGGCCACATGTCGCCAAGAGACACTCGAGAGAGCGAGCAAGACCA 133  
 35 hTAlAspLeuGlyGlnGlyAspIleValGluAlaIaArgValLysAla 51  
 154 CTGAAGCTTGGGCGCAAAAGGTGTTCATTGAGATGTCAGAGGAGTT 203  
 52 LeuAsnThrGlyAlaValAlaIaIaTyrAlaLeuAspLeuArgLysLuph 68  
 204 TGTGAGAGATTCATCTGGCGGCATCCAGTCCAGGCTGTATGAGG 253  
 68 eValArgAspTyrValPheProMetMetArgSerSerAlaLeuTyrGlu 85  
 254 ACCGCTACCTCTGCGACCTCTCTTGGCAGGCGCTGATCGCGCGCA 303  
 85 LyrTyrTyrLeuLeuGlyThrSerIleAlaArgProLeuIleAlaLys 101  
 304 CAAGTGAATTCGCCCGCGGAGGGGGCGCAAGATATGTCTCCAGCGCG 353  
 102 MetValGluIleAlaGluLysGluGlyAlaValAlaIleSerHisGlyAl 118  
 354 CACAGGAAGGGAGACATCAGGTCCGTTTGAAGTCACTGCTCTGAC 403  
 118 aThrGlyLysGlyAsnAspIleValArgPheGluMetSerAlaTyrAla 135  
 404 TGGCGCCCGCATAAAGTGCATCTGCTCCGAGAGATGCTGATTTAC 453  
 135 eulysProAspIleValThrValAlaProThrPargAspTyrPasp..... 149  
 454 AACCGGTTCAAGGCGCGCAATGATGATGATGATGATGATGATGATG 503  
 150 .....PheGlnGlyArgAlaAspLeuGluAlaPheAlaIaArgLysGlu 164  
 504 GATTCCATTCGCGGTACCTCCCAAGACCGGTGAGCATGATGAGAAC 553  
 164 ylleProValProThrThrLysLysAspProThrSerMetAspAlaAsn 181  
 554 TCATGCATCATGATCAGAGGCTGGAATCTGGAACCCCAAGACCA 603  
 181 etLeuHisIleSerTyrGluGlyGlyProLeuGluAspProThrThr 197  
 604 GGGCCCTCCAGTCTACACGAAGACCCAGACCCCAAGCCCA 653  
 198 ProProThrHisMetPheLysLeuThrValAsnProGluAspAlaPro 214  
 654 CAACCCGTGACATTCGAGATCGAGTCAAAAGAGGGCTCTGTAGAG 703  
 214 rGluAlaGluTyrValGluIleGlyTyrValAsnGlyAspProThrSer 231  
 704 TGACCAAGCTAAGAGTGGACACACCCAGACCATCTGAGACCTTC 753  
 231 leAsnGly.....GluGlnLeuSerProAlaIaLeuLeu 242  
 754 ATGTACTGACGAAGTGGCGGCAAGATGCTGGCGCTGTTTACAT 803  
 243 ThrLysAlaAsnGluIleAlaGlyArgHisGlyValGlyValGlyLeu 259  
 804 CTGAGAGACCGCTTCAATGTAATGAGTCCGAGGTATCTACGAGACCC 853  
 259 uValGluAsnArgPheValGlyMetLysSerArgGlyValTyrGluThr 276  
 854 CAGCAGGACCATTCCTTACCATGCTCATTTAGACATGAGGCTTCACC 903

```

|||||.....:|||||.....
276 rogllygthrlleuylrhlslalargalalvalgluserleuthr 292
904 ATGACCGGGAGATGCGCAAAATCAACAGCGCGCTTGAATTTGC 953
293 LeuAspArgIleValLeuHisGlnArgAspAlaLeuGlyProLysTyrAl 309
954 TGAGCGGTATATACCGTTTACGGCTTACCGCTTACCGCTTGA... 996
309 agluLeuValTyrAsnGlyPheThrPheAlaProGluArgGluAlaLeuG 326
997 .....TTTGTCCGCACTGCATCGCAAGTCCAGAGCGAGTGGAAAGG 1041
326 InValTyrPheAspHis...ValAlaLysSer.....ValThrGly 338
1042 AAATGCAAGGTTCCTGCTTCAAGGGCGGCTGATACCTCCGCGCGGA 1091
339 ThrAlaArgLeuLysLeuTyrLysGlyAsnCysAlleValAlaGlyArg 355
1092 GTCCCACTGTCTCTTACAATGAGAGCTGTGAGCATGACGTGCAGG 1141
355 SALAGluArgSerLeuTyrAspLysAspLeuValSerPheGluAlaGly 372
1142 GTGATTATGACCAACATGATGCCACCGGTTTCATCAATCATTCCTTC 1191
372 LysAspTyrAsnGlnHisAspAlaGlyAlaPheIleLysLeuAsnSerLeu 388
1192 AGCGTAGAGAAATATCATCTGTCTCCAGAGCAAG 1224
389 ArgMetArgValGlnLysArgValGluAspLys 399
seq_name: plr2:C83204
seq documentation block:
arginosuccinate lyase PA3525 (imported) - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83204
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lam,
; Lory, S.; Olson, M.V.
N:nture 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337
A:Accession: C83204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-405 <STO>
A:Cross-references: GB:AE004773; GB:AE004091; NID:9949672; PIDN:AAG06913.1; GSPDB:GN003
A:Experimental source: strain PA01
C:gene: argG; PA3525
C:Superfamily: arginosuccinate synthase
alignment_scores:
Quality: 927.50 Length: 404
Ratio: 3.011 Gaps: 6
Percent Similarity: 76.238 Percent Identity: 47.030
alignment_block:
us-09-775-693-1 x C83204
Align seg 1/1 to: C83204 from: 1 to: 405
1 ATGTCCAGCAAGGCTCCGTGCTTACGCTGAGCGGCTGACAC 50
|||||.....
1 MetaAspValLysLysValValLeuAlaTyrSerIleGlyLeuAspTh 17
51 CTCGTCATCTCGTGTGCTGAAGAA...CAAGGCTATGACCTCATTTG 97
|||||.....
17 rserValIleLeuLysTrpLeuGlnAspThrTyrAsnCysGluValValT 34

```

```

98 CCTATCGCCCAACATTGGCCAGAGAGACTTGCAGGAAGCCAGAGAG 147
34 hrPheThrAlaAspLeuGlnGlyGlnGluValAlaGluProAlaArgAla 50
148 AAGGCACTGAGCTTGGGCGCAAAAGTGTTCATTGAGAGATGATGAG 197
51 LysAlaArgAlaMetGluValLysGluIleTyrLleAspAspLeuArgL 67
198 GGAGTTTGGAGAGAGTTCATCTGCGCGGCAATCCAGTCCAGCGCAT 247
67 uGluPheValAlaArgAspPheValTyrProMetPheArgAlaAsnThrValT 84
248 ATGAGCAAGCTTACCTCTGCGCACTCTTTCGAGCGGCTGATCGCC 297
84 YrGluGlnGluTyrLeuLeuGlyThrSerLleAlaArgProLeuIleAla 100
298 CGCAACCAAGTGGAAATCGCGCGGAGCGGCGCAATGATGATGTGCG 347
101 LysArgLeuIleGluIleAlaAsnGluThrGlyAlaAspAlaIleSerH 117
348 CGGCGCCACAGAAAGGGAAGCATGATGATGCTCGGTTGAGCTGATG 397
117 SGlyAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuGlyAla 134
398 ACTCAGTGGCCCCAGATTAAGTCAATGCTGCTCCGAGATGAGCTGAA 447
134 YrAlaLeuLysProGluValLysValIleAlaProThrPargLutPasp 150
448 TTCTTCACACCGGTTTCACAGGCGGCAATGATGATGATGATGATG 497
151 LeuLeuSerArgLys.....LeuMetAspTyrAlaGluLys 163
498 ACACGGGATTCCTCCG.....GTCACGCCAAGAACCCGCGAGCA 541
163 SHSgLyIleProIleGluArgHisGlyLysLysSerProTyrSerH 180
542 TGGATGAGAACTCATGACATGACATGACATGAGGCTGGAATCTGAGA 591
180 etAspLysAsnLeuLeuHisIleSerTyrGlnGlyGlyValLeuGlnAsp 196
592 CCAGAACCAAGCGGCTCCAGGCTGTCTACACGAGAACCCAGACCCAG 641
197 ThrTrpThrGlnHisGlnGluAspMetTrpLysTrpThrAlaSerPro 213
642 CAAGCCCCCAACCCCGATTCGATTCGATTCGATTCGATTCGATTCG 691
213 uAsnAlaProAspThrProThrTyrIleGluLeuThrTyrArgLysGly 230
692 TCCCTGTGAAGTGAACCAAGTCAAGATGAGCAACCCAGACCTCC 741
230 splleValAlaIle.....AspGly...LysAspMetThrPro 241
742 TTGGAGCTTTCATGTACCTGAGCAAGTCCGCGGCAACGATGCGTGG 791
242 AlagluValLeuThrGluLeuAsnArgValGlyGlyIleAsnGlyIle 258
792 CCGTATGACATGCTGAGAACCGCTTCATTGAGATGAAGTCCAGAGTA 841
258 YrArgLeuAspIleValGluAsnArgTyrValGlyMetLysSerArgLys 275
842 TCTAGAGACCCAGCAGGACGATCTTACATGCTCATTTAGACATC 891
275 YrTyrGluThrProGlyGlyThrIleMetLeuLysAlaHisArgAlaIle 291
892 GAGGCTTACCATGACCGGAGAGTCCGCAAAATCAACAGCGCTGG 941
292 GluSerIleThrLeuAspArgGluValAlaHisLeuLysAspLeuLeu 308
942 CTGGAATTTGCTGAGCTGTGTATACCGTTTACGCGCTTACGCGCTG 991
308 TrProLysTyrAlaSerLeuIleTyrThrGlyTyrTrpPserProGlu 325
992 GTGATTTTCCGCGCATGCTCCCAAGTCCAGAGCGAGATGGAAGG 1041

```

```
seq_documentation_block:
  argininosuccinate synthase (EC 6.3.4.5) Cj0665c [Imported] - Campylobacter jejuni (strain
  C:Species: Campylobacter jejuni
  C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
  C:Accession: C81415
```

R.F. Krumholz, J. Wien, B.W. Mungall, K.J. Kelley, J.M. Churcher, C. Basham, D. Chailin, C.W. Quail, M. Rajandream, M.A. Rutherford, K.M. VanVleet, A. Whitehead, S. Barrer, N. Nature 403, 665-668, 2000

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hy

A/Reference number: A01250; M01D:20150912

A/Accession: GCA15

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1406 <PAR>  
A:Cross-references: GB:AL139075; GB:AL11168; NID:969678417; P1DN:CAM75297.1; PID:9696812  
A:Experimental source: serotype O2, strain NCTC 1168  
A:Genetics:  
A:Gene: argG; Cj0665c  
C:Superfamily: argininosuccinate synthase  
;Keywords: ligase

```
alignment_scores:
  Quality: 893.00
  Ratio: 3.014
  Percent Similarity: 74.559
  Length: 397
  Gaps: 6
  Percent Identity: 47.355
```

alignment\_block:

Align seg 1/1 to: C81415 from: 1 to: 406

19 GTGGTCGGCGCTTACACTGGCGGCGTGAACACCTGTGTCATCCCTGCTG 68  
 |||||  
 8 ValValLeuAlaIyrSerIrgylgYLeuAspIrrSerIleIleLeuIstr 24  
 |||||  
 69 GCTGAAG...GAACAAGCGTATGACGTCTATCCATGCTGGCAACATTG 115  
 |||||  
 24 pLeuGlnAspGIuTyrAsnGysGluValValIrrhrrhrrAlaSpIleG 41  
 |||||  
 116 GCCGAGAAGCAAGCTTCGAGGAAGCAGGAAGAGCACTGACCTTGGG 165  
 |||||  
 41 IyGInGlyGluGluLeuGluIrrProAlaIrrGlyIysAlaLeuSerIeNGly 57  
 |||||  
 166 GCCAAAAG.....GTGTTCAATTGAGGATTCGACGAGGAGTTTGGGA 209  
 |||||  
 58 IleIySeluGlnIuAsnIlePheIleIyLysAspLeuArGspIuPheVally 74  
 |||||  
 210 GGAGTTCATCATCGGCGGCATTCACATCCAGGCGCATGTAGGACGCCCT 259  
 |||||  
 74 sAspTyrValIrrPheProMeIrrPheArGAlaAsnAlaIrrIyGlyGluIrrGluT 91  
 |||||

260 ACCTCCCTGGGACACTCTCTCTGGCAGCCCTGCATGCGCCGCAAAACAAGTG 309  
91 yrleuLeuEngIythrInserIleAlaIArgProLeuIleAlaIyIThrGlnAla 1070  
310 GAATCGCCACGCCGAGGGGGCCCAAGATGATGTCTCCACGGCGCCACAGG 359  
108 GlnIleIleAlaEngIInThrIlyAlaIAspAlaIValSerHisIglyAlaThrGln 124  
360 AAGGGGGAACGATACAGTCCGGTTTGAGCTCAGCTGACTACACTGGCCC 409  
124 yLysIglyIAsnAspGlnValAlaArgPhgGlnIleuGlnIYrIleuAlaIAspSer 141  
410 CCAGATTAAGGTGCATGTGCTCCCTGGAGAGATGCTGATTCTCAACACGG 459  
141 roAspIleuIyIleIleIleAlaProIrrPAArgGluTTPAspIleuAsnSerArg 157  
460 TTCAGAGGCCCGCAATGACCTTATGAGTACAGCAACGACAAACGAGATTC 509  
158 GlnIyIyS.....LeuIleuAlaTyAlaIleGlnIyHisIstIyIleAs 170

```

510 CATTCCGGTCACTGCC...AAGAACCCCTGGAGCATGATGAAACSTCA 550
170 pIleSerLySlySLySgLySLySSerProLyZSerMetSLaSLaSLaLeuL 187
557 TGCACATCAGCTTACGAGGTGGAAATCTTGGAAACCCCAAGAACCAAGC 600
187 euhSLleSerLygLygLyLeuValLeuGLuSLaSProlaSLaSLaPro 207
607 CCTCCAGGCTCTACACAGAACCCAGGACCCAGCCAAAGCCCAACAC 656
204 GLUGLuASpMetPrLPrgrtPrSerLySSerProLySLaSLaProSLngl 228
657 CCCCTGACATCTCGAGATCGATGATGATGAAAAAGGGCTCCCTTGAAGTGA 706
220 uSerGLuLeIleGLuLeuASpPrheGLuSLyASpLeuValaLaLea 237
707 CCAAGCTCAAGGATGATGSCAACCCACCCAGACCTCTTGAACCTCTTATG 756
237 SngLygLydSLyS.....LeuSerProLaGLyLeuLeuThr 248
757 TACSTGAACGAAGTGGCGGGACAGCATGGCTGGCCCTATTCAGATCGT 806
249 LysLeuASngLyLeuGLyCySLySLngLyLeGLyLrgrLeuASpLLeVa 265
807 GGAGAACCCGCTTCATGGAAATGAGAGTCCCGAGGATCTACGAAACCCAG 856
265 lGLuSLaSLrgrLyLaLeGLyMetLySSerLaSLrgrLyCySLyGLuThrProG 282
857 CAGGACCATCTCTTACCAGTCTCATTTAGACATCGACGCTTCACCATG 906
282 LygLyThrLeLeuLeuLySLaSLaSLaSLrgrLaLeuGLuSerLeuThrLeu 298
907 GAGCCGGAAGGGCCAAATTCAAACAAAGGCTGTGGCTTAAATTTGCTGA 956
299 ASprrgrGLuLaLaSLaSLaSLeLySLaSPrGLuLeuMeLProLySLyLaSLa 315
957 GCTGGTGTATACCGGTTCAGGCTCAACCCCTGAGCTGTAATTTGGCCGCC 1006
315 rLeuLeLyASngLyTLyTLrPrheSerProSLaSLrgrMeSLeSLeGLuLa 332
1007 ACTGCATCGCCCAAGTCCCAAGAGCAGTGGAAAGGAAATGCGAGGTGCC 1056
332 LaLeuLeASprrgrSLnglLeSLaSLaSLaSLngLyLrgrValLySLeugLy 348
1057 GTCSTCAAGGCGGACGGTGAACATCCCTCGGGCCGGAGTCCCACTG...TC 1103
349 LeuTLyLySLyASnVaLaMeValLeIleGLyLrgrGLuSerLaSLaSLaSPSe 365
1104 TCTTCAATAGAGAGCTGTGAGCATGAAAGCGTGAAGGCTGATATATAGC 1153
365 rLeuPrheASnLaLaLyTLyCySLThrPrheGLuSLuSLyValLySLaSLng 382
1154 CAACGTATGTCACCGGGTTCATCAACATCAATTCCTCTAGG 1194

```

382 InLysAspAlaIaGlyPheIleLysLeuAsnAlaLeuArg 395  
 seq\_name: p1r2:C84048

seq\_documentation\_block:  
 argininosuccinate synthase (citrulline-aspurate ligase) argG [imported] - Bacillus halodurans  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 08-Dec-2000  
 C:Accession: C84048  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai, T.  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; MUID:20263314  
 A:Accession: C84048  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-409 <STO>  
 A:Cross-references: GB:AP001518; GB:BA000004; NID:q10175792; PIDN:BA006906.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: argG  
 C:Superfamily: argininosuccinate synthase

alignment\_scores:  
 Quality: 865.00 Length: 410  
 Ratio: 2.912 Caps: 3  
 Percent Similarity: 72.439 Percent Identity: 44.878

alignment\_block:  
 US-09-775-693-1 x C84048 ..

Align seg 1/1 to: C84048 from: 1 to: 409

```

7  ACCAAGAGCTCCGCTGCTTCCTACAGTGGCGGCTGACACCCGCTG 56
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
2  SerLysLysValValLeuAlaTyrSerLysLysLeuAspHisPheVal 18
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 CATCTCGTGTGCTGAAGAGCAAGGCTATGACGTCATTCGCTATCG 106
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
18  lAlaIleLysTrpLeuSerAspLysGlyTyrAspValIleAlaValGly 35
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 CCACATTTGGCCAGAGCAAGACTTCGAGAGACCCAGAGAGAGCACTG 156
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
35  euAspValGlyGluLysAspLeuGluPheValLysGluValAlaLeu 51
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 AAGCTTGGGGCCMAAAGGTGTTTCATGAGATGTCAGCAGGAGTTCGT 206
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
52  LysValGlyAlaIleGluSerTyrThrIleAspAlaLysLysGluPheAl 68
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 GGAAGACTTCATCTGGCCGGCCATCGACGCGCAGGCTATGAGAGACC 256
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
68  aGluGluPheValLeuProAlaLeuGlnAlaHisAlaLeuTyrGluGln 85
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 GCTACTCTCTGGGACCTCTTTCACAGGCGCTGCATGCCCGCAACAA 306
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
85  ySyrProLeuValSerAlaLeuSerArgProLeuIleSerLysLysLeu 101
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
307 GTGGAATTCGCCAGCGGAGCGGCGCAAGTATGTGCCACGCGGCCAC 356
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
102  ValGluIleAlaGluGlnThrGlyAlaGlnAlaValAlaHisGlyCysTh 118
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 AGGAAGGGAAGCATCAGTCCGTTTGAGTCACTGCTACTCTGCTG 406
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
118  rGlyLysGlyAsnAspGlnValArgPheGluValSerIleGlnAlaLeuAl 135
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
407 CCCCCAGATAAGTCACTTCCTCCGAGGATGAGCTGATTCGATCAAC 456
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
135  snProAsnLeuGluValLeuAlaProValArgGluTrpAlaTrp..... 149
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 CGTTCAAGGCGGCAATGACCTGATGAGTACCAAGCAACACGCGAT 506
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

150  .....SerArgAspGluGluIleGluTyrAlaLysLysAsnAsnI 163
507  TCCCATCCCGGTCACCTCCCAAGAACCCGTCGAGCATGATGAGAACCTCA 556
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
163  eProIleProIleAspLeuAspAsnProTyrSerValAspGlnAsnLeu 180
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
557  TGCACATCAGCTGAGAGCTGGAATCTGGAGAACCCCAAGAACCAAGCG 606
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
180  rPglYArgSerAsnGlyCysGlyIleLeuGluAspProTyrAlaThrPro 196
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
607  CCTCCAGCTCTTACACAGAGAACCCAGACCCAGCAACCAACCCCAACAC 656
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
197  ProGluIleValAlaTyrGluLeuThrValAlaIleGluAspAlaProAsp 213
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
657  CCCGACATTCCTGAGATCGAGTTCAAAAAGGGCTCCCTGTAAGGTGA 706
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
213  nProGluIleValGluIleGluPheGlyLysGlyIleProValThrLeu 229
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
707  CCACGTCACAGATGGCACCCACCCACCAAGCCCTTGAGAGCTTTCATG 756
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
230  ..AsnGlyLysSerTyrProValHis.....GluLeuIleLeu 241
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
757  TACCTGACAGAGATCGCGGCAAGCATGCGTGGCCGTATTCATTCGT 806
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
242  GluLeuAsnGlnIleAlaGlyLysHisGlyValGlyArgIleAspHisVal 258
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
807  GGAGAACCGCTTCATTCGATGAGATGCCGCGATCTACAGACCCAG 856
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
258  lGluAsnArgLeuValGlyIleLysSerArgGluValTyrGluCysPro 275
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
857  CAGGCAACCTCCTTACCATGCTGCTATTCAGATCGAGCCCTTCACCAT 906
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
275  LysAlaMetThrLeuIleLysAlaHisLysGluLeuGluAspLeuThrLeu 291
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
907  GACCGGGAAGTGGCAAAATCAACAAGCCCTTGGAATTCGTCGA 956
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
292  ThrLysGluValAlaHisPheLysProValValGluLysIleAlaG 308
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
957  GCTGGGTATACCGGTTTACGCGCTAGCCCTGAGTGAATTCGTCGCC 1006
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
308  uLeuIleTyrGluLysLeuTyrPheSerProLeuGlnProAlaLeuSerA 325
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1007  ACTGCATCGCCAAAGTCCAGAGAGCGAGTGAAGGAAAGTCAAGAGTGC 1056
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
325  lApeLeuLysGluTyrGlnSerThrValThrGlyValValArgValLys 341
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1057  GTCTCAAGGCGCGAGTGTACATCTCGCGCGGAGTCCCATCTCTCT 1106
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
342  LeuPheLysGlyHisAlaIleValGluGlyArgLysSerGluTyrSerLe 358
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1107  CTACATGAGAGCTGTGTGAGCATGAAAGCGAGCGGTGATTCAGACCA 1156
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
358  uTyrAsnGluLysLeuAlaThrTyrThrProAspAspGluPheAspHisA 375
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1157  CTGATCGCACCGGTTTCATCAATCAATTCCTTCAGGCTGAAGCAATAT 1206
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
375  snAlaIleValGlyPheIleSerLeuTyrPglLeuProThrLysValTyr 391
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
1207  CATCTCTCCAGAGCAAGTCACTGCTCCAAA 1236
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
392  SerMetValAsnLysGluMetLysGluLys 401
   ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq\_documentation\_block:  
 argininosuccinate synthase argG - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: B69589  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C:Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997



A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galled  
 tech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koeltner, P.; Konigstein, G.; Kropp, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Josato, V.; Uchiyama,  
 T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A6580; MUID:98044033  
 A:Accession: B69589  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1,403 <KUN>  
 A:Cross-References: GB:299118; GB:299119; GB:AL009126; MID:92635411; PIDN:CABI4923.1; PI  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: argG  
 C:Superfamily: argininosuccinate synthase

alignment\_scores:  
 Quality: 818.00 Length: 400  
 Ratio: 2.773 Gaps: 3  
 Percent Similarity: 73.750 Percent Identity: 42.750  
 alignment\_block:  
 US-09-775-693-1 x B69589 ..  
 Align seg 1/1 to: B69589 from: 1 to: 403

```

1 ATGTCGACGAAGGCTCCGCGGTTGGGCTTACAGTGGCGGCTGGACAC 50
1 Merserclungllyslsvalvalleuvalatylserglylyleuaspth 17
51 CTGCTGATCTCTGCTGCTGAAGAACAGGCTATGACGCTGATGCTT 100
17 rservalalalelystprleunglglngllytralsvalillelac 34
101 ATCTGCGCAACATGTGGCCGAGGAAAGCATCTGAGGAGGAGGAGAG 150
34 yscylseuaspvalislyslucllylsasprleualalphevaldngllyls 50
151 GCATCGACGCTTGGGCGCAAAAGGCTTTCATGAGGATGTCGAGGAG 200
51 Alaleuclvalalgllyalathrasnserlyrvalilleaspralylscl 67
201 GTTGTGGAGGAGTTCATCTGCGGCGCATCCAGTCCAGCGCATGTAT 250
67 uphealclinsprlyralaleuileserleuclinalanlshmetlyt 84
251 AGGACCGGTACCTCTGCGGCACTCTTTCGAGGCGCTGATCGGCGG 300
84 lugalylslytrproleuvalseralaleuaserlgrproleuilealys 100
301 AAACAAGTGAATCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 350
101 lylleuvalalgluilealaglylsalaspvalalaglnalaleanl 117
351 CGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 400
117 ycsythgllyslgylasnasprlvalalrphelilvalalserlile 134
401 CACTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450
134 etleuasnprleuaspvalalalealprovalalarglvaltrglntr 150
451 TACACCGGTTACAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 500
151 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
serarglclungllylueglutryralalalaserat 162

```

```

501 CGGATTTCCCATCCCGGTCTACCTCCCAAGACCCGCTGAGCATGATGA 550
162 gelyliletrproletrleaslnleuaspserprotyrserilleaspr 179
551 ACCATGACATCATGAGTACGAGGCGGAGGAGGAGGAGGAGGAGGAGG 600
179 snleutprglvalalalasnslgyllyleuclualaspptprpala 195
601 CAAGCGCTCCAGCTCTTACACGAGGAGGAGGAGGAGGAGGAGGAGG 650
196 AlaprotroglulvalatylraserleuThrAlprouleuclulstpr 212
651 CAACACCCCTGACATCTGAGATGAGTCAAAAAGGCGCTGATGA 700
212 oasprtrproglulvalilleglulilealrphelclungllyalpr 229
701 AGGTGACCAAGCTCAAGATGCGACACCCAGGAGGAGGAGGAGGAGG 750
229 erlle.....AsprlyalserlyrserleuSer...Glutleu 240
751 TTCATGTACCTGACGAGTCCGCGGCAAGCATGCGTGGCGGATATGA 800
241 lleleuylsleuasnslumelalaglyalalnslyalalargylleas 257
801 CATGCTGAGAACCGCTTATGGAATGAGTCCCGGATGATCGAGAGA 850
257 phlsvalglusnasrlyleuvalalgllyleuaseralglulvalyl 274
851 CCGGACGCGGACCATCTTACCATGCTCATTTAGACATCGAGGCTTC 900
274 ysproglyalalmetrthleuileuylalalnslyslunglunlsp 290
901 ACATGAGACGGGAGAGTCCCAAAATCAAAAGGCGCTGAGCTTAAAT 950
291 Thrleuvalylslylvalalalnsphelysproillelglungllyl 307
951 TGCTGACCTGCTGATACCGGTTTACGCGCTGAGCTGAGTGAATTT 1000
307 tserglulileleryrasnlyleuThrpheserproleuylaspral 324
1001 TCCGACATGATCGGCAAGTCCGAGGAGGAGGAGGAGGAGGAGGAG 1050
324 euhlsalrphelueuylslylthrglnlyshlsvalthlelylval 340
1051 GTGCGCTGCTCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1100
341 Vallyleuupheueylshlsalalevalalglulgllylsserly 357
1101 GTCTCTTCAATGAGGAGTGTGATGATGAGTGTGATGAGTGTAT 1150
357 rserleuyltraspgrlulysleuAlalathrlyllyllylasp 374
1151 AGGCAAGTGAATCGGCGGTTTCATGATCAATTCCTCCAGGAGGAG 1200
374 sprhishlsalalalellylthleuileuclultrpgrleuProth 390

```

seq\_name: p1r2:T06667

seq\_documentation\_synch:  
 argininosuccinate synthase (EC 6.3.4.5) - *Arabidopsis thaliana*  
 N:Alternate names: protein F617.40  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence-revision 23-Apr-1999 #text-change 22-Oct-1999  
 C:Accession: T06667  
 R:Bevan, M.; Terry, N.; Ardiles, W.; Buysmaert, C.; Dasseville, R.; De Clerck, R.;  
 Ohlssel, J.; Mewes, H.W.; Meyer, K.F.X.; Scheller, C.  
 submitted to the Protein Sequence Database, April 1999  
 A:Reference number: 215792  
 A:Accession: T06667  
 A:Molecule type: DNA  
 A:Residues: 1,498 <BEV>  
 A:Cross-References: EMBL:AL049657; GSPDB:GN00062; ATSP:F617.40  
 A:Experimental source: cultivar Columbia; BAC clone F617



C:Genetics:  
 A:Gene: ATSP:6617.40  
 A:Map position: 4  
 A:Introns: 25/1; 100/1; 146/2; 162/3; 223/3; 285/3; 315/2  
 C:Superfamily: argininosuccinate synthase  
 C:Keywords: ligase

alignment\_scores:  
 Quality: 815.00 Length: 396  
 Ratio: 2.890 Gaps: 5  
 Percent Similarity: 71.212 Percent Identity: 42.172

alignment\_block:  
 US-09-775-693-1 x T06667 ..

Align seg 1/1 to: T06667 from: 1 to: 498

```

19 GTGTTCTGGCTTACAGTGGCGGCGGACACCTGCTGCTGCTG 68
   |||
128 ValValLeuAlaTyrSerGlyGlyLeuAspThrSerValIleValProthr 144
   |||
69 GCTGAAGAACAA...GGCTATGACGTATGCTGCTGCTGCTGCTG 115
   |||
144 pLeuLysLysnTyrGlyGlyValValLysPheThrAlaAspValG 161
   |||
116 GCCAGAG...GAGACTTCGAGAGACCCGAGAGAGAGAGAGAGCTT 162
   |||
161 LylGlnGlyLleLysGlnLeuGlnLysGlnLysAlaLysAlaSer 177
   |||
163 GGGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 212
   |||
178 GlyAlaSerGlnLeuValValLysAspLeuThrGlnLysValLys 194
   |||
213 GTTCATCTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 262
   |||
194 pPheLlePheProCysLeuArgAlaGlyAlaLysLysLysLys 211
   |||
263 TCCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 312
   |||
211 euleuGlyThrSerMetAlaArgProValIleAlaLys..... 223
   |||
313 ATGCCCGAGCGGAGGGCGGCAAGTATGTCGCCAGCGGCCACAGGAA 362
   |||
223 ..... 223
   |||
363 GGGGAGCATCAGGTCCGCTTGTGAGTCACTGCTGCTGCTGCTGCTG 412
   |||
224 .....ValArgPheGlnLeuThrPhePheSerLeuAsnProG 236
   |||
413 AGATTAAGCTCATTTGCCCTGCGAGGATGCTGATTTACAAACCGGTT 462
   |||
236 LulLeuLysValValAlaProThrPArgLutProLuu.....Ile 248
   |||
463 AAGGGCGGCAATGACCTGATGAGTACGCAAGACACACGCGGATTCCT 512
   |||
249 GlnGlyArgGlnAspAlaIleGlnLysLysLysLysLysLysValPro 265
   |||
513 CCCGCTACTCCCAAGAACCCCTGAGATGATGATGATGATGATGATG 562
   |||
265 LProValThrLysLysSerLysSerLysSerLysSerLysSerLys 282
   |||
563 TCGAGTACGAGGATGCTGAGAACCCCAAGAACCAAGCCGCTCCA 612
   |||
282 euehrlsGlnGlyAspLeuLysLysLysLysLysLysLysLysLys 298
   |||
613 GGTCTTACACGAGACCCGAGAGCCGAGAGAGAGAGAGAGAGAGAG 662
   |||
299 AspMetLysMetLysSerValAspProGlnAspAlaProAspGln 315
   |||
663 CATTCCTGAGATGAGTTCAAAAAGAGGCTCCCTGTAAGTGAACCA 712
   |||
315 uTyrlleGlnLleGlyLleGlnSerGlyLeuProValAlaLeuAsnGly 331

```

```

713 TCAAGATGCGACACCAACCAACGACCTGCTGAGCTGCTGCTGCTG 762
   |||
332 .....LysAlaLeuSerProAlaThrLeuLeuAlaGlnLeu 343
   |||
763 AACGAGTCCGCGGCAAGCATGCGCTGCGCTGCTGCTGCTGCTGCTG 812
   |||
344 AsnThrIleGlyLysLysLysLysLysLysLysLysLysLysLys 360
   |||
813 CCCTGCTTACGATGAGTACGCTGAGTATCTACGAGACCCGAGAGCA 862
   |||
360 nArgLeuValLysLysSerLysSerLysLysLysLysLysLysLys 377
   |||
863 CCATCCCTTACGATGCTGATTTAGACATGAGGCTTCAACATGAGCG 912
   |||
377 ThrLeuPheAlaAlaValGlnLysLysLysLysLysLysLysLys 393
   |||
913 GAATGCCGCAAAATCAACAGAGGCTGAGTGAATTTGCTGAGCTGT 962
   |||
394 GluSerLleGlnValLysAspThrLeuAlaLeuLysLysLysLys 410
   |||
963 GTATACCGGCTTACGCGCTAGCCTGAGTGTGAATTTGCTGAGCTG 1012
   |||
410 LTyrlAlaGlyArgTyrPheAspProLeuArgLysSerMetLysPAlaPhe 427
   |||
1013 TCGCCAAATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062
   |||
427 eTcLysLysLysLysLysLysLysLysLysLysLysLysLysLys 443
   |||
1063 AAGGCGGAGGTACATCCTCGCGCGGAGTCCGCTGCTGCTGCTGCT 1112
   |||
444 LysLysSerValSerValThrGlyArgLysSerProAsnSerLysTyr 460
   |||
1113 TGAGAGCTGCTGAGTACATGAGCTGAGGCTGATTTATGAGCCACT 1162
   |||
460 gLlnAspLleSerSerPheGlnLysSerGlnLysLysLysLysLys 477
   |||
1163 CCACCGGCTTACATCATCATTCCTCAGGCTGAG 1200
   |||
477 lAlaGlyPheLleArgLeuTyrGlyLeuProMetLys 489

```

seq\_name: p1r2:E84935

seq\_documentation\_block: (EC 6.3.4.5) [Imported] - Buchnera sp. (strain APS)

argininosuccinate synthase (EC 6.3.4.5) [Imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: E84935

R:Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.

A:Reference number: A84930; M01D:20445173

A:Accession: E84935

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-403 <STD>

A:Cross-references: GB:AP000398; GSPDB:GM00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: argG; BU050

C:Superfamily: argininosuccinate synthase

C:Keywords: ligase

alignment\_scores:  
 Quality: 807.50 Length: 396  
 Ratio: 2.710 Gaps: 5  
 Percent Similarity: 75.253 Percent Identity: 43.687

alignment\_block:  
 US-09-775-693-1 x E84935 ..

Align seg 1/1 to: E84935 from: 1 to: 403

```

19 GTGGTTCCTGCTACAGTGGGCGCTGACACCTGCTGATCTCTGTG 68
9 ValValLeuAlaTyrSerGlyLeuAspThrSerAlaIleIleProtr 25
69 GCTGAAGGAGCA... GCTATGACCTGATTCCTATCTGCGCAACATG 115
25 pleuLysGluAsnTyrAsnPhelGluValAlaIleAlaIleAspIleG 42
116 GCCAG... AAGAGACCTTCGAGAGGAGGAGGAGGAGGAGGAGGAG 162
42 LysIleSerLysAspLeuAsnGlyIleGluLysLysSerLeuGlnSer 58
163 GGGGCAAAAGAGTGTTCATTCAGAGTTCAGAGGAGGAGGAGGAGGAG 212
59 GlyAlaSerSerCysHisValPheAspLeuGlnGluIleLeuGln 75
213 GTTCATCTGCGCGGCGGATCCAGTCCAGGCGGATGATGAGGAGGAG 262
75 nTyValTyrProValLeuLysThrGlyAlaLeuTyrGlnGlySerTyr 92
263 TCCTGGGACCTCTCTTCCAGGCGGCTGATCCGCGGCAAGAGGAGAA 312
92 euleGlyThrAlaMetAlaArgProIleAlaLysGlnValGlu 108
313 ATGCGCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 362
109 LeuAlaLeuAsnIleGlyAlaAsnSerLeuGlnGlyAlaThrGly 125
363 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 412
125 sGlyAsnAspIleValArgPheGlnMetAlaTyrAlaIleAlaPro 142
413 AGATTAAGGATCATTCCTCCGAGGAGGAGGAGGAGGAGGAGGAGG 462
142 snLeuAsnValIleAlaProtrParg... GluTrp... AsnLeu 154
463 AAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 512
155 AsnSerArgGlnSerLeuLysLysLysLysLysLysLysLysLys 171
513 CCGGCTACCTCCAGAGACCGGAGGAGGAGGAGGAGGAGGAGGAGG 562
171 rThrAlaThrLeuGlnLysIleTyrSerLysAspGluAsnSerTrpHis 188
563 TCAGCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 612
188 IeSerThrGlnGlyLeuLeuLysProIleAlaIleAlaIleAlaIle 204
613 GGTCTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 662
205 AspCysTrpSerTrpThrValAsnProGluAspAlaProGluLysPro 221
663 CATTCGAGAGTCAAGTCAAAAAGGAGGAGGAGGAGGAGGAGGAGG 712
221 uTyrValSerLeuGlnLysGlnGlyValValSerValAsnAsnG 238
713 TCAAGAGTGACACCCAGACGAGGAGGAGGAGGAGGAGGAGGAGG 762
238 Lys... LeuAsnProLeuLysCysValGlnLeu 249
763 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
250 AsnSerLeuGlyAlaLysHisGlyIleGlyArgIleAspIleLeuG 266
813 CCGCTCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862
266 nArgLeuIleGlyMetLysSerArgGlyCysTyrGlnThrProGly 283
863 CCATCCTTACACGCTCATTCATTCAGAGGAGGAGGAGGAGGAGGAG 912
283 hTleIleMetThrAlaIleLysAlaIleGlnLeuValLeuAspArg 299

```

```

913 GAATGCCCAAAATCAACAGAGGCGCTGCTGAATTTGCTGACTGCT 962
300 GluSerPheArgTrpArgGluLysIleGlyLeuGlnMetSerSerLeu 316
963 GTATACGCGTTTACGCGCTTACGCGCTTACGCGCTTACGCGCTTAC 1012
316 LTyAspGlyArgTrpPheSerProIleArgLysSerLeuGlnAlaIle 333
1013 TCGCCAAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1062
333 LaAspSerLeuSerLeuGlnIleThrGlyLeuValIleLeuLysLeuTyr 349
1063 AAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1112
350 LysGlySerValThrAlaValGlnLysSerProAsnSerLeuTyr 366
1113 TGAGAGCTGCTGAGCATCAACATCAATTCCTCAGGCGGAGGAG 1162
366 rGlnGluTyrAlaThrPheGlyGluAspLysValTyrLysGlnSerAsp 383
1163 CCAGCGGCTTCATCAACATCAATTCCTCAGGCGGAGGAG 1200
383 LaAspGlyPheIleArgLeuPheSerLeuSerLys 395

```

seq\_name: p1r2:B70398

seq\_documentation\_block:

```

arginosuccinate synthase
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:9619666
A:Accession: B70398
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-401 <AOP>
A:Cross-references: GB:AE000725; NID:g2983598; PIDN:AAO7170.1; PID:g2983600; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: argG
C:Superfamily: arginosuccinate synthase

```

alignment\_scores:

```

Quality: 800.50 Length: 404
Ratio: 2.741 Gaps: 3
Percent Similarity: 72.277 Percent Identity: 41.832

```

alignment\_block:

```

US-09-775-693-1 x B70398 ..
Align seg 1/1 to: B70398 from: 1 to: 401

```

```

10 AAGAGCTCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 59
3 LysLysArgValIleLeuAlaTyrSerGlyGlyLeuAspThrSerIleI 19
60 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109
19 eValArgTrpLeuThrGlnGlyGlyTyrGlnValIleThrLysThrAla 36
110 ACATTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 159
36 sPValGlyGlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 52
160 CTGAGGCGGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 209
53 AlaGlyAlaIleGlnAlaIleValGlnAspLeuLysGlnThrPheAlaG 69

```

Wed Feb 13 07:36:06 2002

us-09-775-693-1.rpt

page 15

```
210 GGAGTTCATCTGCGCGGCATCCAGTCCAGCGACGTCGATGAGACCGCT 259
||||| |||||||
69 uasnTyrCysLeuProThrLeuAlaLeuAlaLeuTyrGluGlyLysT 86
|||||
260 ACCTCTGGGACACCTCTTCGCGAGCCCTGATCGCCGCGCAACAGAGG 309
|| ||| |||||||
86 TyrProLeuThrAlaAlaLeuSerAlaProLeuIleAlaGluAlaGluVal 102
|||||
310 GAATCCCGCCAGCGGAGGCGCCAGATGTCGTCACCGCGCCAGCAG 359
|||||
103 TyrTyrAlaGluLysPheAsnAlaAspTyrAlaAlaHisGlySerThrG 119
|||||
360 AAAGGAGGAGATCAGTCCGCTTTCAGTCTGCTGCTACTGCTGCTG 409
|||||
119 LysGlyAsnAspGlnValAlaArgPheGluLeuSerValTyrAlaLeuAsn 136
|||||
410 CCCAGATAAGCTCATGCTCTCCGAGATGCTGAAATTCACACCGG 459
|||||
136 roAspIleGluValLeuAlaProValAlaArgGluTyrPglu..... 148
|||||
460 TTCAGAGCGCGCATGATGATGATGATGATGATGATGATGATGATG 509
|||||
149 PheLysSerArgGluGluGluGluGluGluGluGluGluGluGlu 559
|||||
510 CATCCCGGTCACCTCCCAAGAACCCGTCGAGCATGATGATGATGAT 182
|||||
165 ovalLysAlaThrLysGluLysProTyrSerIleAspArgAsnLeuTyr 609
|||||
560 ACATCAGCTACGAGCTGATGATGATGATGATGATGATGATGATGAT 198
|||||
182 LysAlaSerIleGluLysGlyProLeuGluAspProTyrGluGluPro 215
|||||
610 CCAGTCTCTACAGAGACGAGACGAGACGAGACGAGACGAGACGAG 230
|||||
199 GlnAspAlaTyrGlnIleThrGlnSerProGluGluAlaProAspGlu 245
|||||
660 TGACATCTCGAGATCGAGTCCAAAGAGGCTCCCTGTGAAGTGACCA 290
|||||
215 ogLysValIleThrValIleGluPheGluLysGlyLysProValTyrLeu... 230
|||||
710 ACGTCAAGATGCGACACCCAGCCAGACCTCTGAGCTCTCATGTAC 259
|||||
231 .....AsnGlyGluArgTyrGluGluGlnTyrPylsLeuIleAlaAsn 244
|||||
760 CTGACCAAGTCCGCGGCAAGCATGCGCTGCGCTGATGATGATGATG 809
|||||
245 LeuAsnGluIleAlaGluArgHisGlyValGlyArgIleAspMetValG 261
|||||
810 GAACGCGTTCATGGAATGATGATGATGATGATGATGATGATGATG 859
|||||
261 uasnArgLeuValGlyIleLysSerArgGluIleTyrGluAlaProGly 278
|||||
860 GCACCATCTTACCATGCTCATTTAGACATGACGAGCCCTCCACATG 909
|| ||| |||||||
278 IawaValLeuTyrGluAlaTyrArgAspLeuSerLeuValLeuAsp 294
|||||
910 CGG...GAAGTCCGCAAAATCAACAAGAGGCTGCGCTGAAATTTGCTGA 956
|||||
295 ArgPheThrPheHisTyrPheLeuThrHisIleProHisGlyTyrAlaL 311
|||||
957 GCTGGTGTATACGGTTTACGGCTGAGCCCTGAGTGTGATTTGTCGCC 1006
|||||
311 sLeuValTyrGluGlyLeuTyrPheThrProLeuArgGluAlaLeuAsp 1056
|||||
1007 ACTGCATCGCCAGTCCAGAGCGAGTGAAGGAAAGTCAAGTCAAGT 1106
|||||
328 IapHerhAsnLysIleAlaGluPheAlaThrGlyGluValAlaArgLeu 344
|||||
1057 GTCTCAAGGCGCAGGTGTATCTCTCGCGCGAGTCCCGCATGTCTCT 1106
|||||
345 LeuTyrLysGlySerValSerValAlaGlyArgArgSerProAsnSer 361
|||||
1107 CTGCAATGAGGAGTGTGAGCATGAAGTGCAGGGTGTATGAGCCAA 1156
|||||
```

Wed Feb 13 07:36:06 2002

us-09-775-693-1.rpt

Page 16

OM of: US-09-775-693-1 to: SwissProt\_39.\* out\_format: pfs  
Date: Feb 12, 2002 4:11 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODE=frame+asp.model -DEV=xlp  
-O/cgn2.1/USPTO.spool/US09775693/runat\_12022002.124154.1491/app-query.fasta.1.1315  
-DB=SwissProt\_39 -OPMT=fastan -SUFFIX=esp -CAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -XGAPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct  
-THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NOR=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09775693\_0CGN1\_1\_36 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPRX  
-WAIT -THREADS=1

## Search information block:

Query: US-09-775-693-1  
Query length: 1239  
Database: SwissProt\_39.\*  
Database sequences: 100059  
Database length: 3664827  
Search time (sec): 71.550000

score\_list:  
Sequence Strid Orig ZScore EScore Len Documentation  
SwissProt\_39:ASSY\_HUMAN + 2144.00 2323.43 2.3e-122 412 P09034 rattus norvegicus (rat)  
SwissProt\_39:ASSY\_RAT + 2086.00 2260.62 7.4e-119 412 P16460 mus musculus (mouse)  
SwissProt\_39:ASSY\_MOUSE + 2073.00 2246.54 4.5e-118 412 P14568 bos taurus (bovine)  
SwissProt\_39:ASSY\_BOVIN + 2069.00 2242.21 7.8e-118 419 P09069 drosophila melanogaster  
SwissProt\_39:ASSY\_DROME + 1164.00 1172.05 3.0e-63 410 P09354 schizosaccharomyces  
SwissProt\_39:ASSY\_SCHPO + 1083.00 1164.51 2.3e-58 420 P22768 saccharomyces cerevisiae  
SwissProt\_39:ASSY\_YEAST + 1069.00 1159.16 1.6e-57 420 P09523 arabidopsis thaliana  
SwissProt\_39:ASSY\_ARATH + 911.50 986.80 5.2e-48 523 P09347 bacillus subtilis  
SwissProt\_39:ASSY\_BACSU + 818.00 887.69 2.2e-42 403 P03437 bacillus subtilis  
SwissProt\_39:ASSY\_BACAT + 807.50 876.32 9.6e-42 403 P17138 buchera aphidicola  
SwissProt\_39:ASSY\_BOCAL + 800.50 868.78 2.5e-41 401 P07213 aquifex aeolicus  
SwissProt\_39:ASSY\_AQUAE + 777.00 829.37 6.6e-40 398 P57793 lactococcus lactis  
SwissProt\_39:ASSY\_LACTA + 766.00 829.37 4.0e-39 400 P77973 synecococcus sp. (st  
SwissProt\_39:ASSY\_STINY + 750.00 814.86 2.6e-38 390 P08032 methanococcus jannasch  
SwissProt\_39:ASSY\_METHA + 747.00 810.97 4.3e-38 395 P060174 methanococcus jannasch  
SwissProt\_39:ASSY\_METVA + 731.00 793.00 3.9e-37 397 P13256 methanococcus vanniel  
SwissProt\_39:ASSY\_METHU + 692.50 751.87 8.3e-35 399 P13252 methanobacterium thet  
SwissProt\_39:ASSY\_METTH + 686.00 744.85 2.0e-34 398 P04993 mycobacterium tuberco  
SwissProt\_39:ASSY\_METRA + 686.50 741.10 3.3e-34 396 P13257 methanococcus barkeri  
SwissProt\_39:ASSY\_METRA + 670.50 728.08 1.8e-33 397 P50986 streptomyces clavulig  
SwissProt\_39:ASSY\_METRA + 630.50 684.60 4.6e-31 405 P08176 corynebacterium glut  
SwissProt\_39:ASSY\_METRA + 621.00 664.11 6.4e-30 401 P57877 pasteurella multocida  
SwissProt\_39:ASSY\_METRA + 621.00 664.11 6.4e-30 401 P57877 pasteurella multocida  
SwissProt\_39:ASSY\_PASMO + 421.50 456.94 2.0e-18 445 P57877 pasteurella multocida  
SwissProt\_39:ASSY\_PASMO + 413.00 448.32 6.0e-18 444 P43313 haemophilus influenza  
SwissProt\_39:ASSY\_HAELI + 391.00 424.46 1.3e-16 446 P22767 escherichia coli  
SwissProt\_39:ASSY\_HAELI + 386.00 418.42 4.6e-16 446 P06734 streptomyces lavendul  
SwissProt\_39:ASSY\_STRPA + 333.00 381.05 4.1e-13 480 P24532 streptomyces coelicol  
SwissProt\_39:ASSY\_STRPA + 150.00 146.47 0.0493 3530 P10162 homo sapiens (human)  
SwissProt\_39:PPPL\_HUMAN + 144.00 160.93 0.0987 276 P03181 epstein-barr virus (str  
SwissProt\_39:PPPL\_HUMAN + 142.50 152.14 0.1275 660 P03181 epstein-barr virus (str  
SwissProt\_39:CM44\_HUMAN + 141.50 143.32 0.1543 1690 P53420 homo sapiens (human)  
SwissProt\_39:DBPL\_RAT + 139.50 144.09 0.1997 1183 P54258 rattus norvegicus (rat)  
SwissProt\_39:DBPL\_HUMAN + 139.00 143.53 0.2141 1185 P54259 homo sapiens (human)  
SwissProt\_39:GLTA\_HUMAN + 138.50 143.84 0.2252 838 P08489 tritium aestivum (wt  
SwissProt\_39:PPPM\_HUMAN + 136.00 153.62 0.2972 234 P10161 homo sapiens (human)  
SwissProt\_39:PPPM\_HUMAN + 135.00 150.77 0.3029 331 P04288 homo sapiens (human)  
SwissProt\_39:PPPL\_HUMAN + 135.00 144.56 0.3372 660 P08488 tritium aestivum (wt  
SwissProt\_39:GLTO\_WHEAT + 134.50 143.62 0.3870 648 P10387 tritium aestivum (wt  
SwissProt\_39:MWIS\_MOUSE + 134.50 129.73 0.4244 3511 P09424 tritium musculus (mouse)  
SwissProt\_39:ASSY\_LACLC + 131.50 159.80 0.5160 61 P05491 lactococcus lactis (su

seq\_name: SwissProt\_39:ASSY\_HUMAN  
seq\_documentation\_block:  
ID ASSY\_HUMAN STANDARD; PRT; 412 AA.  
AC P09066; 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DR 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE  
LIGASE).  
GN ASS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84015388; PubMed=6194510;  
RA Bock H.-G.O., Su T.-S., O'Brien W.E., Beaudet A.L.;  
RT "Sequence for human argininosuccinate synthetase cDNA."  
RL Nucleic Acids Res. 11:6505-6512(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84135824; PubMed=6321498;  
RA Freytag S.O., Bock H.-G.O., Beaudet A.L., O'Brien W.E.;  
RT "Molecular structures of human argininosuccinate synthetase  
pseudogenes. Evolutionary and mechanistic implications."  
RL J. Biol. Chem. 259:3160-3166(1984).  
RN [3]  
RP SEQUENCE OF 200-209.  
RX TISSUE-Colon carcinoma;  
RA MEDLINE=97295306; PubMed=9150948;  
RJ H.L., Reid G.E., Moritz R.L., Edges J.S., Burgess A.W., Simpson R.J.;  
RT "A two-dimensional gel database of human colon carcinoma proteins."  
RL Electrophoresis 18:605-613(1997).  
RN [4]  
RP VARIANTS S-14; H-157; N-180; S-324; W-304; W-363 AND R-390.  
RX MEDLINE=90293089; PubMed=2358466;  
RA Kobayashi K., Jackson M.J., Tick D.B., O'Brien W.E., Beaudet A.L.;  
RT "Heterogeneity of mutations in argininosuccinate synthetase causing  
human citrullinemia."  
RL J. Biol. Chem. 265:11361-11367(1990).  
RN [5]  
RP VARIANTS LEU-18 AND CYS-86.  
RX MEDLINE=92048472; PubMed=1943692;  
RA Kobayashi K., Rosenblum C., Beaudet A.L., O'Brien W.E.;  
RT "Additional mutations in argininosuccinate synthetase causing  
citrullinemia."  
RL Mol. Biol. Med. 8:95-100(1991).  
RN [6]  
RP VARIANTS THR-118; VAL-192; CYS-272; ARG-280; TRP-304 AND LEU-363.  
RX MEDLINE=95067972; PubMed=797368;  
RA Kobayashi K., Shabehn N., Terazono H., Saheki T.;  
RT "Mutations in argininosuccinate synthetase mRNA of Japanese patients,  
causing classical citrullinemia."  
RL Am. J. Hum. Genet. 55:1103-1112(1994).  
RN [7]  
RP CHARACTERIZATION OF SOME VARIANTS.  
RX MEDLINE=96385010; PubMed=8792870;  
RA Shabehn N., Kobayashi K., Terazono H., Fukushige T., Horikuchi M.,  
Saheki T.;  
RT "Characterization of human wild-type and mutant argininosuccinate  
synthetase proteins expressed in bacterial cells."  
RL Enzyme Protein 48:251-264(1995).  
RN [8]  
RP CATALYTIC ACTIVITY: AMP + L-CITRULLINE + L-ASPARTATE = AMP +  
PYROPHOSPHATE + L-ARGININOSUCCINATE.  
RN [9]  
RP -I- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC

CC PATHWAY.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- DISEASE: DEFECTS IN ASS ARE THE CAUSE OF CITRULLINEMIA, AN  
 CC AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SEVERE VOMITING  
 CC SPELLS AND MENTAL RETARDATION.  
 CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X01630; CAA25771.1; -.  
 DR EMBL; L00084; AAS1783.1; -.  
 DR EMBL; L00079; AAS1783.1; JOINED.  
 DR EMBL; L00080; AAS1783.1; JOINED.  
 DR EMBL; L00081; AAS1783.1; JOINED.  
 DR EMBL; L00082; AAS1783.1; JOINED.  
 DR EMBL; L00083; AAS1783.1; JOINED.  
 DR PIR; A01195; AJHURS.  
 DR MIM; 603470; -.  
 DR MIM; 215700; -.  
 DR InterPro; IPR001518; Arginosuc\_synth.  
 DR Pfam; PF00764; Arginosuc\_synth; 1.  
 DR ProDom; PD003544; Arginosuc\_synth; 1.  
 DR PROSITE; PS00564; ARGININOSUCCIN\_SYN\_1; 1.  
 DR PROSITE; PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
 KW Arginine biosynthesis; urea cycle; Ligase; ATP-binding;  
 KW Disease mutation.  
 FT NP\_BIND 115 123  
 FT VARIANT 14 14  
 FT VARIANT 18 18  
 FT VARIANT 86 86  
 FT VARIANT 118 118  
 FT VARIANT 157 157  
 FT VARIANT 180 180  
 FT VARIANT 192 192  
 FT VARIANT 272 272  
 FT VARIANT 280 280  
 FT VARIANT 304 304  
 FT VARIANT 324 324  
 FT VARIANT 363 363  
 FT VARIANT 363 363  
 FT VARIANT 390 390  
 FT SEQUENCE 412 AA; 46426 MW; D9083E0655D0716 CRC64;

alignment\_scores:  
 Quality: 2144.00  
 Ratio: 5.204  
 Percent Similarity: 100.000  
 Length: 412  
 Gaps: 0  
 Percent Identity: 100.000

alignment\_block:  
 US-09-775-693-1 x ASSY\_HUMAN ..  
 Align seg 1/1 to: ASSY\_HUMAN from: 1 to: 412

1 ATGTCAGCAAGAGCTCCGTGTTCCGGCTCAGTGGCGGCTTGAGAC 50  
 1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspThr 17  
 51 CTGCTGACCTCTCGTGTGGCTGAAGAACCAAGCTATACGTCATTCGCT 100  
 17 rSerCysIleLeuValTyrLeuLysGluGlnClyTyrAspValIleAlaT 34  
 101 ATCTGGCCAACTTGGCCAGGAAGAACTTCGAGGAAGCCAGGAAGAG 150  
 34 YrLeuAlaAsnIleLysGlnLysGlnAspPheGluAlaIleArgLysLys 50  
 151 GCATGAAGCTTGGGGCCAAAAGGTGTTCATTGAGATGTCCACAGAGA 200  
 51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerAlaGly 67  
 201 GTTTGTGAGAGAGTTCATCTGGCCGGCCATCCAGTCCAGCGCACTGATG 250  
 67 uPheValGluGluPheIleTyrProAlaIleGlnSerSerAlaLeuTyrG 84  
 251 AGAACCGCTACCTCCTGGGACACTCTCTGGCCAGCGCTGCATGCGCCG 300  
 84 LuAspArgTyrLeuLeuGluTyrSerLeuAlaIleArgProCysIleAlaArg 100  
 301 AAACAAGTGGAAATCCGCCAGCGGAGGGGCCAAGTATGTGTCCACGG 350  
 101 LysGlnValGluIleAlaGlnArgGluGlyAlaLysTyrValSerHisGln 117  
 351 CGCCACAGGAAGGAGGAGCATCGTCGCGTTTGACTGACTGACTACT 400  
 117 YalaThrGlyLysGlyAsnAspGlnValArgPheGluLeuSerCysTyrS 134  
 401 CACTGGCCCCCAGATAAAGTCATTGCTCCCTCGAGAGATGCTGAATTC 450  
 134 erLeuAlaProGlnIleLysValIleAlaProTyrPheMetProGluPhe 150  
 451 TACAACGGTTCAAAGGCCCAATGACCTGATGAGTACGCAAGAACA 500  
 151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnH 167  
 501 CGGAGATTCCATCCCGGTCACTCCCAAGAAGCCGTGAGATGATGATGAGA 550  
 167 sGlyIleProIleProValThrProLysAsnProTyrPheMetAspGluA 184  
 551 ACCTCATGCAATCAGCTACGAGAGCTGGAATCCTGAGAACCCCAAGAAC 600  
 184 snLeuMetHisIleSerTyrGluAlaGlyTyrLeuGluAsnProLysAsn 200  
 601 CAAGCCCTCCAGGCTCTACAGGAAGACCCAGGACCCCAAGACCCC 650  
 201 GlnAlaIleProIleGluTyrThrLysTyrGlnAspProAlaLysAlaBr 217  
 651 CAACACCCCTGACATTCAGATGAGTCAAAAAGGGGCTCCCTGAGA 700  
 217 oAsnThrProAspIleLeuGluIleGluPheLysLysGlyAlaProValL 234  
 701 AGGTGACCAAGCTCAAGATGCGACACCCACAGACCTCTTGGAGGCTC 750  
 234 yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGluLeu 250  
 751 TTCATGTACCTGAAGAGTCCGGGCAAGCAATGCGCGTGGCCGATATGA 800  
 251 PheMetTyrLeuAsnGluValAlaGlyLysHisGlyValAlaArgTyrLeu 267  
 801 CATGCTGAGAACCGCTTCATTTGATGAATGAAGTCCGAGGTATCTACGAGA 850  
 267 rIleValGluAsnArgPheIleGlyMetLysSerArgGlyIleTyrGluT 284  
 851 CCCAGCAGGAGCAATCTTACCATGCTCAATTAAGACATCGAGGCTTC 900  
 284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300

Wed Feb 13 07:36:07 2002

us-09-775-693-1.15P

seq\_name: SwissProt\_39:ASSY\_RAT  
seq\_documentation\_block:  
ID ASSY\_RAT STRAND: PRT: 412 AA.  
AC P09034; 1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, last sequence update)  
DT 01-NOV-1988 (Rel. 09, last annotation update)  
DT 15-JUL-1999 (Rel. 38, last annotation update)  
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE  
DE LIGASE).  
GN ASS-  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney; PubMed=3174461;  
RX MEDLINE=89016648; Pubmed=3174461;  
RA Such L.C., Morris S.M., O'Brien W.E., Beaudet A.L.;  
RT "Nucleotide sequence of the cDNA encoding the rat argininosuccinate  
RT synthetase";  
RL Nucleic Acids Res. 16:9352-9352(1988).  
CC -1- PROPHOSPHATE + L-ARGININOSUCCINATE  
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: X12459; CAA30999.1; -  
CC EMBL: M36708; AAA0771.1; -  
CC PIR: S01440; AURRS.  
CC InterPro: IPR001518; Argininosuccinyl-  
CC Pfam: PF00764; Argininosuccinyl-1.  
CC ProDom: PD003544; Argininosuccinyl-1.  
CC PROSITE: PS00564; ARGININOSUCCIN\_SYN\_1; 1.

DR PROSITE: PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
KW Arginine biosynthesis; Urea cycle; Ligase; ATP-binding.  
FT NP\_BIND 115 123 ATP (POTENTIAL).  
SQ SEQUENCE 412 AA; 46496 MW; CCA80906F5A3E93D CRC64;

alignment\_scores: Length: 412  
Quality: 2086.00 Gaps: 0  
Ratio: 5.113  
Percent Similarity: 99.029 Percent Identity: 96.845

alignment\_block:  
US-09-775-693-1 x ASSY\_RAT

Align seg 1/1 to: ASSY\_RAT from: 1 to: 412

1 ATGTCGACCAAGAGGCTCGGTTCCTGAGTACAGTGGCGCTGAGAC 50  
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspPh 17  
51 CTGTCGATCTCTCTGCTGCTGAGAGCAAGAGCTATGACGTATTCCT 100  
17 rSerCysIleLeuValIlePheLysGlyGlnGlyTyrAspValIleAla 34  
101 ATCTGCGCAACATGCGCCAGAGAGACCTTCAGAGAGCCAGAGAGAG 150  
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaIleArgLys 200  
151 GCACGTAGAGCTTGGGGCAAAAGGTTTCATTCATTCAGATGTCAGAGGA 200  
51 AlaLeuLysIleGlnAlaLysValPheIleGlnAspValSerLysGln 67  
201 GTTGTGGAGAGTTCATCTGCGCGCATTCGACGTCAGCCAGCCACTGATG 250  
67 upPheValGlnGlnPheIleTyrProAlaValGlnSerSerAlaLeuTyr 84  
251 AGCAGCGTACTCTCTGCGACCTCTCTGCGACGCTGACGCTGACGCGCC 300  
84 IuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100  
301 AAACAGGTGAAATGCGCCAGCGGGAGGGGCAAGTATGTCCTCCAGCG 117  
101 LysGlnValGlnIleAlaGlnIleArgGlnIleLysTyrValSerHisGln 117  
351 CGCCAGCGAAAGAGGAGAGATCAGTCCGCTGAGTTCAGTTCAGTTCAG 134  
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuThrCysTyr 134  
401 CACTGCGCCCGACATTAAGTCATTCCTGCGAGAGATGCTGAGATTC 450  
134 eTleuAlaProGlnIleLysValIleAlaProTyrPArgMetPheGlnPhe 150  
451 TACAACCGGTTCAGAGGCGCCAGTACGCTGATGAGATGATACCAAGACA 500  
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGlnHis 167  
501 CGGATTCCTCCATCCCGGTACCTCCCAAGAACCGGTGAGAGATGATGAGA 550  
167 sGlyIlePheIleProValThrProLysSerProTyrSerMetLeuAspGln 184  
551 ACCATGACATCAGCTACGAGGTGAGTATCTGAGAACCCCAAGAAC 600  
184 sLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200  
601 CAAGGCGCTCAGGTCTACAGCAAGAACCCCAAGAACCCCAAGAACCC 650  
201 GlnAlaArgProGlnLysTyrThrLysTyrGlnAspProAlaLysAla 217  
651 CAACACCGCTGACATTCGAGATCGAGTTCAGAGAGGGTCCCTGAGGA 700  
217 oAsnThrProAspValLeuGlnIleGlnPheLysGlyValProVal 234  
701 AGGTGACCAACTCAAGATGCGACACCCAGCCAGCAGCTCTTGGAGCTC 750

```

|||||
234 yvalThrasValLysaspGlyThrHisSerThrSerLeuAspLeu 250
751 TTCATGTACCTGACAGAGTCCGGGCAAGCATGCGTGGCCGTATATGA 800
251 PheMetTyrLeuAsnGluValAlaGlyHisGlyValGlyArgLeuAs 267
801 CATCGTGGAGAACCGCTTCATGGATGAGTCCCGAGGTATACGAGA 850
267 pIleValGluAsnArgPheIleGlyMetTyrSerArgGlyTyrGlu 284
851 CCCGACGACGACCATCTTACCATGCTCATTTAGACATCGAGCCCTC 900
284 htrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAla 300
901 ACCATGAGCCGCAAGTCCGCAAAATCAACAGCCCTGGCTGAATTT 950
301 ThrMetAspArgGluValAlaArgTyrIleGlyGlnGlyLeuGlySph 317
951 TGCTGAGCTGGTATACCGGTTTACGCGCTAGCCCTGAGTGTGAATTG 1000
317 eAlaGluLeuValTyrThrGlyPheThrPheHisSerProGlyCysGlu 334
1001 TCCGCACTGCAATGCCCAAGTCCGAGAGGAGTGAAGGAAAGTGCAG 1050
334 aArgHisCysTyrIleAspTyrSerGlnGluValGluGlyValGln 350
1051 GTGCCGCTCTCAAGGCGCAGGTGTACATCTCGCGCGGAGTCCCACT 1100
351 ValSerValPheTyrGlyGlnValTyrIleLeuGlyArgGluSerPro 367
1101 GTCTCTACATGAGAGTGGTGGATGATGAGTGCAGGATGATATG 1150
367 uSerLeuTyrAsnGluGluLeuValSerMetAsnValGlnGlySph 384
1151 AGCCACTATGTCACCGGCTTCATCAATCAATCCCGTCCAGCTGAG 1200
384 luoProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeu 400
1201 GAATATCATGCTCTCAAGCAGCAAGTGCATGCCCAA 1236
401 GluTyrHisArgLeuGlnSerTyrValThrAlaLys 412

seq_name: SwissProt_39: ASSY_MOUSE

seq_documentation_block:
ID ASSY_MOUSE STRAND: PRT: 412 AA.
AC P16460:
DT 01-AUG-1990 (rel. 15, Created)
DT 15-AUG-1990 (rel. 15, Last sequence update)
DE "Molecular characterization of the murine argininosuccinate
DE LIGASE".
DE ASS OR ASS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBL_taxid=10090;
RP [1]
SEQUENCE FROM N.A.
RC STRAIN=DA/2J; TISSUE=Liver;
RA MEDLINE=91216457; PubMed=1708740;
RT Surh L.C., Beaudet A.L., O'Brien W.E.;
RT "Molecular characterization of the murine argininosuccinate
RT synthetase locus.";
RL Gene 99:181-189(1991).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: M31690; AAA37266.1; -
DR EMBL: M31692; AAB60707.1; -
DR EMBL: M31694; AAB60708.1; -
DR EMBL: M31693; AAB60708.1; JOINED.
DR EMBL: M31695; AAB60708.1; JOINED.
DR PIR: M31702; AAB60706.1; -
DR MGD: MGI:88090; Ass1.
DR InterPro: IPR001518; Argininosuc_synth.
DR Pfam: PF00764; Argininosuc_synth.1.
DR ProDom: PD003544; Argininosuc_synth.1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
KW Arginine biosynthesis; Urea cycle; Ligase; ATP-binding.
FT NP_BIND 115 123 ATP (POTENTIAL).
SQ SEQUENCE 412 AA; 46584 MW; A8F3AFDDFBAF6A CRC64;

alignment_scores:
Quality: 2073.00 Length: 412
Percent Similarity: 98.786 Gaps: 0
US-09-775-693-1 x ASSY_MOUSE Percent Identity: 96.117

alignment_block:
Align seg 1/1 to: ASSY_MOUSE from: 1 to: 412
1 ATGTCCAGCAAGAGCTCCGTGTTGCGCTACAGTGGCGGCTGAGAC 50
1 MetSerSerTyrGlySerValAlaLeuAlaTyrSerGlyGlyLeuAspThr 17
51 CTGCTGATCTCTCGTGGCTGAGAGCAACAGCATGATGCTATGCTCT 100
17 rSerCysIleLeuValAlaThrPheLeuGlnGlyValAlaPheValIleLeu 34
101 ATTCGGCCAAATGCGCAGAGAGACTTCGAGGAAGCCAGCAAGAAAG 150
34 TyrLeuAlaAsnIleGlyGlnTyrGlnAspPheGlnGluAlaArgTyr 50
151 GCACGTGAAGCTTGGGGCAAAAGGTCTTATTGAGAGTGCAGAGGA 200
51 AlaLeuLysLeuGlyAlaLysValAlaPheIleGluAspValSerTyr 67
201 GTTGTGAGAGAGTTCATGCTGCGCGCCATCCAGTCCAGCGCAGCTATG 250
67 uPheValGluGlnuPheIleThrProAlaValGlnSerSerAlaLeuTyrG 84
251 AGGACCGCTACCTCTCGGACCTGCTTGCAGAGCGCTGATGCCCGC 300
84 luoAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 100
301 AAACAAGTGAATCCCGCAGCGGAGGGCCCAAGTATGTGTCCACGG 350
101 ArgGlnValGluIleAlaGlnArgGluGlyAlaLysTyrValSerHis 117
351 CGCAGCAGGAAGGGAGCAATGATGAGTCCGTTGAGCCATGATGCTACT 400
117 yAlaThrGlyTyrGlySerAsnAspGlnValAlaArgPheGluLeuThrCysTyr 134
401 CACTGCGCCCGCCAGATAAAGTCAATGCTCCCGGAGAGATGCTGAATT 450
134 eTleuAlaArgGlnIleTyrValIleAlaProIlePheTyrMetProGluPhe 150
451 TACAACCGTTCAAGGCGCAATGACTGATGAGTACGCAAGCAACA 500

```



```

|||||
151 TyAsnArgPheLysGlyValGlnAspMetGluTyrAlaLysGlnI 167
501 CGGGATTCCATCCCGGTCCTCCAGAACCCGCGAGCATGGATGAGA 550
167 sglYlleProIleProValThrProLysSerProTrpSerMetLspglua 184
551 ACCTGATGACATCATGAGTACGAGGCGGGAATCCTGAGAACCCAGAAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAACGGCTCCAGGCTCTTACAGGAAGCCAGACCCGCGCAAGGCC 650
201 GlnAlaProProGlyLeuTyrTyrIleGlnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATCTCTGAGATCGAGTTCAAAAAAGGGCTCTGTGA 700
217 oAsnSerProAspValLeuGlnIleGlnPheLysGlyValProValL 234
701 AGTGATGACAGTCAAGATGACACCCAGAACCCGCGAGCATGGATGAGA 750
234 yValThrAsnIleLysAspGlyThrThrArgThrTrpSerLeuGlnLeu 250
751 TTCATGATACCTGAGAGTCCGCGGCAAGCATGCGGCGCTATGGA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIleAs 267
801 CATGATGAGAACCCGCTTATGATGATGATGATGATGATGATGATGATG 850
267 PileValGlnAsnArgPheIleLysMetLysSerArgIleTyrGluT 284
851 CCCGAGAGGACACATCCTTACATGATGATGATGATGATGATGATGATG 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 300
901 ACCATGACCGGAGTGGCCAAATCAAAACAAAGAGCTGAGCTGAAT 950
301 ThrMetAspArgGlnValArgLysIleLysGlnIleLysGlyLeuLysPh 317
951 TGCTGAGTGGTGTATACCGGTTTACCGGCTTACCGGCTTACCGGCTT 1000
317 eAlaGlnLeuValIleThrGlyPheThrPheTrpHisSerProGlnLysGlnPhe 334
1001 TCCGCCACTGATCCGCAAGTCCCAAGGAGGAGTGAAGGAAAGTGCAG 1050
334 aAlaArgHisCysIleGlnLysSerGlnGlnArgValGlnGlyLysValGln 350
1051 GTGTCCGCTCTCAAGGCGGAGGTGATACCTCTGCGCGGAGTCCCACT 1100
351 ValSerValPheLysGlyIleValIleLysGlyArgGlnSerProLe 367
1101 GTCTCTCTACATGAGAGTGGTGTGAGCATGAGTGAAGGAGTGAAT 1150
367 uSerLeuTyrAsnGlnGlnLeuValSerMetLsnValGlnGlyAspTyrG 384
1151 AGCCAATGATGCCACCGGCTTATCATCAATCATCAATCATCAATCAT 1200
384 lUpProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTCTCCAGAGACGATCAGTCCAA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: SwissProt_39: ASSY_BOVIN
seq_documentation_block:
ID ASSY_BOVIN STANDARD; PRT; 412 AA.
AC P14568;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LIGASE).

```

```

GN ASS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046714; PubMed=2813370;
RA Dennis J.A., Healy P.J., Beaudet A.L., O'Brien W.E.;
RT "Molecular definition of bovine argininosuccinate synthetase
RT deficiency."
RT Proc. Natl. Acad. Sci. U.S.A. 86:7947-7951(1989).
RL -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- DISEASE: DEFECTS IN ASS ARE THE CAUSE OF A BOVINE FORM OF
CC CITRULLINEMIA.
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M26198; AAA30388.1;
DR PIR: A33986; AUBORS.
DR InterPro: IPR001518; Arginosuc_synth.
DR Pfam: PF00764; Arginosuc_synth; 1.
DR ProDom: PD003544; Arginosuc_synth; 1.
DR PROSITE: PS00564; ARGININOSUCCIN SYN.1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN SYN.2; 1.
DR Arginine biosynthesis; urea cycle; Ligase; ATP binding.
KW NP_BIND 115 123
FT SEQUENCE 412 AA; 46417 MW; 6F7AC7F445EED086 CRC64;
SO

```

```

alignment_scores:
Quality: 2069.00 Length: 412
Ratio: 5.059 Gaps: 0
Percent Similarity: 99.272 Percent Identity: 95.388

```

alignment\_block:

us-09-775-693-1 x ASSY\_BOVIN ..

Align seg 1/1 to: ASSY\_BOVIN from: 1 to: 412

```

1 ATGTCACGAAAGCTCCGTTGCTGCTACAGTGGCGGCTGGACAC 50
|||||
1 MetSerGlyLysGlySerValValLeuAlaLysSerGlyGlyLeuAspPh 17
51 CTCGTGATCTCTGCTGTGCTGAGAGAACAAAGCTATGAGCTGATGCT 100
17 rSerGlyIleLeuValTrpLeuLysGlnGlnGlyTyrAspValIleAla 34
101 ATCTGGCCACACTGTGGCCAGAAAGAACTTCGAGAGCCAGAGACAG 150
|||||
34 yLeuAlaAsnIleGlyLysGlnAspPheGlnGlnAlaIleArgLysLys 50
151 GCACGTAAGCTTGGGGCCAAAGAGTGTTCATGTAGAGATGCACAGGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValAlaPheIleGlnAspIleSerLys 67
201 GTTTGTGAGAGACTTCATCTGGCGGCGCATCCAGTCCAGGCACTGATG 250
|||||
67 uPheValGlnGlnPheIleTrpProAlaIleGlnSerSerAlaLeuTyrG 84
251 AGAACCGCTACTCTCTGGAGCACTCTTGGCCAGGCGCTGCATGCCCGC 300

```

```
384 luprVAlAspaIatHrcgYPheILeasnlLeasnSerleuaTArgLeuLys 400
```

1201 GAATCATCAGTCGCTCCAGAAGACGTCACTGCCAA 1236  
|||||  
**401** GIUTYTHSArGLEuglnslValtThraLaLys **412**

seq\_name: SwissProt\_39\_porc ...

seq\_name: SwissProt 39.1ecv brows

seq\_documentation\_block:

AC 097069; Q9VIA1: PRT; 419 AA.

DT 30-MAY-2000 (Rel. 39, Created)

20-AUG-2001 (Rel. 40, Last annotation update)

ASPARTATE LIGASE) (EC 6.3.4.5) (CITRULLINE-...

*Drosophila melanogaster* (Fruit fly)

Platyzoa; Metazoa; Arthropoda; Tracheata; Hexanoda; Insecta

Ephydroidea; Drosophilidae; Diptera; Brachycera; Muscomorpha;

[1]

STRAIN=BERKELEY.

Palazzo M.T. S.E., Pte

complete sequence of the Antennapedia complex of *Drosophila melanogaster* (Submitted 14-11-2004)

[2] EMBL/GenBank/DBJ databases.

STRAIN-BERKELEY:

Adams M D Colnaghi P  
MEDLINE=20196006; PubMed=10731132;

Amannalides P.G., S.  
Georgas D.

Sutton G.G., Wortman

Wan K.H., Doyle C.

Ballew R.M. Basil

DEESON K.Y., Benos  
Borkova D. 2011

Burtis K.C., Busam

de Pablos B., Delch

Durbin K.T. Euraco

Foster C., Gabrieli  
Clodagh A.

Harris N.L., Harvey

Jalali M., Kalush F

Lasko P. E., kodlra

LIU X., Mattei B., N  
Merklon

Mount S.-M., Moy M.,

Palazzo M., Pittma

Shue B. C.

Spreading

Wang Z.-Y., Wassarma.

Ye J., Yeh R.-F., Zhai

Gibbs R. A. Murray F. I.

The genome sequence

-!- CATALYTIC ACTIVITY

### THE FIVE POINTS

\_\_\_\_\_

# THE FINAL STEP OF THE ARGinine BIOSYNTHETIC

```

alignment_scores:      length:      405
Quality:      1164.00      Gaps:      2
Ratio:      3.571
Percent Similarity:      80.494      Percent Identity:      54.321

```

1 lysglutthvalilleuallleuallisercllyglyleuasphtsercysva 15  
 3 lysglutthvalilleuallleuallisercllyglyleuasphtsercysva 109  
 60 ccctcgctgctgtcaaggaacagcctatgacgtctatccctatctggcca 36  
 19 lleuylstptlelleuasphtlysglntgcyglvalllecyvalleuallaa 159  
 110 acattggcccaaggaagaaacattctcgaggaagccaggaacagcactgagc 52  
 36 spvaltglyglntlysglasphtthralaalglwlysglyslalleuay 209

[illegible][illegible]

alignment\_scores:

Quality:	1083.00	Length:	395
Ratio:	3.384	Gaps:	4
Percent Similarity:	81.013	Percent Identity:	54.430

US-09-775-693-1 x ASSY\_SCHPO

to: ASSY\_SCHPO from: 1 to: 410

```

42 GTTGGCCATCAAGTGGCGCCCTGGACACCTCGTCATCCGTGCTGCT
9 ValLeuAlaTyrSerIleTyrLeuAspThrSerCysIleLeuAlaTrpLe
72 GAAGCAAAAGGCTATGACGTATTGCTCATCTGCGCCAAACATTGGCCAGA
1 uileGluGluGlyTTPGluValIleIecysTyrMetAlaAsnValGlyGluG
122 AGCAACACTTCGAGCAACCCAGGAAGAAAGGCACACTGAAGCTGGGGCCAA
42 ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
42 InGluAspTrpAspAlaIaIaTrpGluIuLysAlaLeuGlyValGlyAlaLys
172 AAGGTCTCATTTGAGCATGTCACACGAGGAGTTGTGGAGAGACTCATCTG
59 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
59 LysValTyrValGluAspLeuArgGluGluIuPheIleAsnAspTrpValIle
222 GCCGCCATCCAGTCCAGCGCACCTGATGATGAGACCGCATCCCTCGTGACA
75 |||||:::|||||:::|||||:::|||||:::|||||:::|||||
75 ePrOAlaIaIaGlnAlaAsnAlaIleTyrGluAsnValTyrLeuLeuGlyT
272 CCTCTCTTCGACAGCCCTGCATCGCCCGCAAAACAAGGCAAAATCGCCGAC
92 hrSerLeuAlaIaTrpIleIleAlaIaGlnArgGlnIleGlnIleAlaGlu
322 CGGAGGGGGCCAGTATGTGTCCACGCGCCACAGGAAGGGGAAACA 371
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

109 LysGIuSnCyStIleAlaValSerHisGlyCysThrGlyLysGlyAsnAs 125  
372 TCAGGTCGGTTTGGAGTCAGCTGCTACTCAGCTCAGTCCGCCCCAGATAAG 421  
125 pGIuValArgpHeGluLeuAlaTyrTyrAlaLeuLysProAspValGln 142  
422 TCATTGCTCCCGGAGGATGGCTGAAATTGTATCAACAGCGTTCAAGGGCCG 471  
142 alIleAlaProThrParGluLeuProValPhepHeGluArgpHealGlyArg 158  
472 AATGACCGTAGAGAGCTACGCAAAAGCAACAGGGATGCCATCCGCTAC 521  
159 LysAspLeuLeuGlnTyrAlaAlaValLysGlyIleProValThrGlnThr 175  
522 TCCCAAGACCCGCTGAGCATGATGAGACCTCATGACATTCAGTAC 571  
175 rThLysLysProTsrPserMetAspLysAsnIleValHisCysSerTyrG 192  
572 AGCGTGAATCCCGGAGAACCCCAAGAACAGAGCGCTCCAGGTCCTAC 621  
192 IuAlaGlyIleLeuGlnLysProSerMetThrProProLysAspMetTrp 628  
622 ACAGAACCCAGAGACCCACAGCCAAAGCCCAACACCCCTGACATTCGCA 671  
209 LysLeuThrValAspProLysAspAlaProAspGlnValGln 672  
672 GATTCAGTTCAAAAAGCGCTCCCTGGAAGCGTACGACCAACGTCAAGATG 721  
225 rIleHisPheGluLysGlyAlaProThrLysLeu...GluCysLysAspG 241  
722 GCACCCACCCACACCTCTCTGGAGCTTTCATGTCATTCAGTCAAGAGTC 771  
241 Lys...ThrPserGlyValAlaSerIlePheTyrGlnLeuAlaIle 772  
772 GCGGCGAAGACGGCGCCGATGATGACATCTGGAGAACCGCTTCAT 821  
257 AlaArgArgAsnGlyValGlyArgIleAspIleValGluAsnArgPheSe 872  
822 TGGAAAGAAGTCCGAGGTATCTAGACAGACCCACAGGACACCATCTTT 871  
273 rGlyLysSerSerArgLysCysTyrGlnThrProGlyLeuThrIleLeuA 290  
872 ACCATGCTCATTTAGACATGAGAGCGCTTACCAATGACCGGGAAGTGGC 921  
290 rGlnrAlaHisMetAspLeuGlnGlyLeuThrMetGlnArgIuValArg 968  
922 AAAATCAAA...CAAGCGCTGGGCTTGAATTTGGTACGGGTATAC 968  
307 AlaLeuArgAspGlnPheValThrPheAsnLeuAlaLysIleLeuLys 968  
969 CGGTTTAAAGCCTAAGCCTGAGTGTGATTTGTCGGCCACTGGATCGCA 1018  
323 nGlyGlnPhePheSerProCysThrArgMetLeuAlaAlaAsnAsn 1018  
1019 AGTCCAGAGAGCAATGGAAGGAAATGCAAGGTGCTGCTTCAAGGCG 1068  
340 alSerGlnGlnValValAsnGlyValAlaLysLeuSerValTyrLysGly 1068  
1069 CAGGTGTCATTCCTCGGCGGAGAGTCCCACTGCTC...CTCTACATGA 1115  
357 AsnValThrValLeuGlnTyrGlyLysSerAspThrAlaHisLeuTyrAspG 1115  
1116 GGACCTGTCAGCATGAACGTGACAGGAGATTAATGAGCAACATGATGCA 1165  
373 uLysLeuSerSerMetAspGlnLeuGlnGlyLysPheAspProThrTrpThrS 1166  
CGGGTTCATCAACATCAATTCCTCAGGCTGAAG 1200  
390 eGlyPheIleGlnIleLeuSerMetArgLeuArg 1401

seq\_documentation block

```

alignment_scores:      Length:  414
                       Quality: 1069.00
                       Ratio:   3.351
                       Percent Similarity: 77.053
                       Percent Identity: 50.483

alignment_block:
  us-09-775-693-1 x ASST_YEAST ..
  Align seg 1/1 to: ASST_YEAST from: 1 to: 420

157 AAGCTTGCGCCAAAAGGNGTTCATGTGAGATGTGCACGAGGATTGT 206
152 lyslllelglyalacysylsphenavalcyvalaspcysargluaspheva 68
207 GGAAGAGTTCATCTGGCGCGGCATCCAGTCCAGCGCAGCTATGAGGACC 256
68 llyaspilleuupheproalvalalgnvalnsalavalvalyrgluaspy 85
257 GCTACTCTCTGGGACCTCTCTTGCCAGGCCCTGCATCGCCCGCAACAA 306
85 allyleuleuenglythrsleuallarpvoalllealalysalagln 101
307 GTGGAAATGCCCCAGCGGAGGGGCCAAGATGTGTCCACGCGCCAC 356
102 lilespalalalysglnglnllyshelalalserhislscysrh 118
357 AGGAAAGGGGAACGATCAGTCCGGTTTGAGCTAGCTGCCTACTCACTG 406
118 rgllygelylashnspglnllearpheglndeusephethyllalaleu 135
407 CCCCCAGATAAGGTCATGTGCCCTCGAGAGATCCCTGAATTTCTACAC 456
135 ysprowaspvalilyscysilrheprrotprargmetprogluupheglu 151
457 CGGTCGAAGGGCCCAATGACTGTAGTGAAGTACCAAGAACCAACGGGAT 506
152 AtgphelalaglyargylasplendeuasptryalalaglnlyglY11 168
507 TCCCATCCCGGTCACCTCCACAGAACCCGTGAGCATGATGAGAACCTCA 556
168 eprovalalaglnthrlysalalysprotpserthrAspGlunAsglna 185
557 TGCACATACGACTACGAGGCTGGAATTCCTGGAGAAACCCACAGAACG 606
185 lahlslleserlyrglnalaglylleasgluasprowasprthrhrpro 201
607 CCTCAGCTGTCTACACAGAAACCCAGAGACCCAGCCAAAGCCCAACAC 656
202 Prolysaspmetirpysleuilevalaspprometaspalalaspog1 218
657 CCCTGCATCTTCGACATGCGATTCGAAAAGGGGTCCCTGTGAAGTGA 706
218 nroglinspmetthrllleasphgelturrgelyleuprovalylsleut 235
707 CCAAGGTCAAGATGTGACCAACCCACAGACTCC.....TTG 744
235 hrrythrAspalslythrserylslnvalservalthrlysprou 251
745 GAGCTCTTCACTACCTGACGAGAGTGGCGGCAACATGCGGTGGCG 799
:::.....:::

```

```

252 AspvAlpheLeuAlaAlaSerAsnLeuAlaAlaGlyAlaGlyAr 268
795 TATTGACATGTCGAGACCGCTTCATTGGAATGAATGCCAGGATATCT 844
268 gilleAspIleValGluAspArgTyrIleAsnLeuLysSerArgGlyCysT 285
845 ACAGAGACCCAGCAGACACCATCTTCATGCTCATTTAGCATCGAG 894
285 yrlcIuAlaIaProLeuThrValLeuAlaArgLysAlaHisValAspLeuGlu 301
895 GCCTTCACCATGACCGGAGAGTCGCAAAATCAACACAGGC...CTGGG 941
302 GlyLeuThrLeuAspLysGluValAlaGlnLeuAlaArgAspSerPheValTh 318
942 CTGGAATTTGTCGACGCTGCTGATACCGCTTACGGCTACCGCTGAGT 991
318 ProAsnThrSerArgLeuIleTyrAsnGlyPheLeuLeuHisProGluC 335
992 GTGAAATTTGTCGCCCATGCTGACCGCAGCTGACGAGCAGGAGAGG 1041
335 ysgIuTyrIleArgSerMetIleGlnProSerGlnAsnSerValAsnGly 351
1042 AAAGCTGAGGTGTCGCTCCTCAAGGCCAGCTGATCATCTCGGCGCGGA 1091
352 ThrValArgValArgLeuTyrGlyAsnValIleIleLeuGlnArgSe 368
1092 GTCCCCACG...TCTCTTCAACATGAGAGAGCTGTGATCATCAACGTC 1138
368 rTrIuTyrThrGluTyrAspProThrGluSerSerMetAspGluL 385
1139 AGCGTGATTTATGACGACCATGATGCCCGGCTTCATCAACATCAATTC 1188
385 euThrGlyPheLeuProThrAspThrIleAlaIleAlaGlnAla 401
1189 CTCAGCTGAGCAATATCATCTCTCCAGACGAGGATCAGT 1230
402 IleArgIleLysTyrGly.....GluSerLysTyrThr 413

seq_name: SwissProt_39:ASSY_ARATH
seq_documentation_block:
ID ASSY_ARATH STANDARD; PRT; 523 AA.
AC Q952X3;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ARGININOSUCCINATE SYNTHASE, CHLOROPLAST PRECURSOR (EC 6.3.4.5)
GN ARG424830 OR F617.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Mambo R., Murphy G., Voickaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansong W., Brandt P., Griwall L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Meche R., Meier M.,
RA Kreis M., Delseny M., Fuldomech P., Watson M., Schmidheini T.,
RA Reihert B., Porteleille D., Perez-Alonso M., Boutry M.,
RA Vos P., Hohelsel U., Zimmermann W., Medler H., Bancroft I.,
RA Langham S.-A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chung Y.-D., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Braun M.,
RA Woolfman P., Klein Lankhorst R., Rose M., Hauf J., Koelke W.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gieten J., Villarroel K., De Clercq R.,

```

```

RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K.,
RA Petre A., Rajandream M.-A., Lyne M., Benes V., Reemann S.,
RA Borkova D., Bloeker H., Scharte M., Grimm M., Loehert T.-H.,
RA Dose S., de Haan M., Maize A., Schaefer M., Muller-Auer S.,
RA Neumann S., Argillon B., Grandpré K., Danner D., Herzl A.,
RA Massenet O., Quigley F., Vitale D., Liguori R., Pivard E.,
RA Schnabl S., Hiller R., Schmidt W., Lechay A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casadhera E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Teyl J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Frisman D., Haase D., Lemcke K., Heber S., Francis P., Bielle C.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Basile M., Habermann K.,
RA Parnell L., Dedbia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Selkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mox P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston K., Yil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shoddy N., Hasegawa A., Rodriguez M., Hoffman J., Till S.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RA thaliana."
RA Nature 402:769-777(1999).
RN
RP CONCEPTUAL TRANSLATION.
RL Lemke C.;
RU
CC -1- UNPUBLISHED OBSERVATIONS (JUN-2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE -> AMP +
CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO WRONG EXON
CC PREDICTIONS FROM THE GENOMIC SEQUENCE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC
DR EMBL; AL049657; CAB41123.1; ALT_SEQ.
DR EMBL; AL161562; CAB79393.1; ALT_SEQ.
DR InterPro; IPR001518; ArgInosuc_synth.
DR Pfam; PF00764; ArgInosuc_synth; 2.
DR ProDom; PD003544; ArgInosuc_synth; 1.
DR PROSITE; PS00564; ARGININOSUCCIN SYN.1; 1.
DR PROSITE; PS00565; ARGININOSUCCIN SYN.2; 1.
KW Arginine biosynthesis; ArgInosuc; Ligase; ATP-binding; Transil peptide;
FT CHLOROPLAST (POTENTIAL).
FT CHAIN 1 523
FT SEQUENCE 523 AA; 57249 MW; 97B82A5A718969C8 CRC64;

```

## alignment\_scores:

```

Quality: 911.50
Ratio: 3.018
Percent Similarity: 76.263
Length: 396
Gaps: 4
Percent Identity: 45.707

```

## alignment\_block:

```

US-09-775-693-1 x ASSY_ARATH

```



Align seg 1/1 to: ASSY\_ARATH from: 1 to: 523

```

19 GTGTTGTCGCTACAGTGGCGCTGACACCTGCGTCTGCTGCG 68
128 ValValLeuLeuValSerGlyLeuValSerValLeuValProtr 144
69 GCTGAAGAACAA... GGTATGACGCTCATTCCTATCTGCGCAACAT 115
144 pLeuValSerGlyLeuValSerGlyLeuValSerGlyLeuValSer 161
116 GCGCAGAG... GAACTTGGAGAACGCGCAAGAACGCACTGAGCTT 162
161 LysGlyLeuValSerGlyLeuValSerGlyLeuValSerGlyLeu 177
163 GGGGCGCAAAAGGTGTCATGAGATGTCACGAGGAGATGTCAGAG 212
178 GYALASerGlnLeuValValLysAspLeuThrGlnLeuValLys 194
213 GTTCATCTGCGCGGCGCATCCAGTCCAGCGCATGATGAGACCGCT 262
194 pPheLeuValSerGlyLeuValSerGlyLeuValSerGlyLeuVal 211
263 TCCTGGGACCTCTCTTGGCAGGCGCTGAGTCCGCGCAACAGTGA 312
211 euleuGlyThrSerMetAlaArgProValIleAlaLysAlaMet 227
313 ATGCGCCAGCGGAGGCGGCGCAATATGTCTCCCGCGCGCGCAAG 362
228 ValAlaAlaGlnValGlyAlaAlaValAlaLysGlyCysThrGly 244
363 GGGGACGATCAGTCCGCTTGGAGCTGAGCTGCTACTGCTGCTG 412
244 sGlyAsnAspGlnValAlaArgPheGlnLeuThrPhePheSer 261
413 AGATTAAGCTATGCTGCTGCTGAGATGCTGCTGCTGCTGCTG 462
261 LLeuValValAlaAlaProTrpArgGlnTrpGln... 273
463 AAGGCGCCCATGACCTGATGATGACGAAAGCAACAGCGGATTC 512
274 GlnGlyArgGlnAspAlaIleGlnValAlaLysLysAsnValPro 290
513 CCGGCTACTCCCAAGAACCGCGTGGAGATGATGAGAACCTCAT 562
290 LProValThrLysLysSerLysLysSerLysLysSerLysLys 307
563 TCAGCTACGAGGCTGATCTGAGAACCGCGTGGAGATGATGAG 612
307 euserHisGlnLysLysSerLysLysSerLysLysSerLysLys 323
613 GGTCTCTACAGAAAGCAAGCGCGTGGAGATGATGAGAACCT 662
324 AspMetLysMetLysValAspProGlnAspAlaProGlnPro 340
663 CATCTCTGATGATGATGATGATGATGATGATGATGATGATG 712
340 uTryIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 762
713 TCAGAGATGCGACACCGACCGCTCTTGGAGATGATGAGAAC 368
357 ..... LysAlaLeuSerProAlaThrLeuLeuAlaGlnLeu 812
763 AAGCAAGTCGCGGCAAGCATGCTGCGCTTATGATGCTGAGAA 812
369 AsnThrIleGlnIleGlnIleGlnIleGlnIleGlnIleGln 385
813 CCGCTTATGATGATGATGATGATGATGATGATGATGATGATG 862
385 mArgLeuValGlyMetLysSerArgGlyValAlaValAlaVal 402
863 CCATCTTATGATGATGATGATGATGATGATGATGATGATG 912

```

```

402 hTrileuPheAlaAlaValGlnGlnLeuGlnSerLeuThrLeuAspArg 418
913 GAACTGCGCAAAATCAACAGCGCTTGAATTCGAGAGCTGCT 962
419 GluSerIleGlnValArgPhePheProLeuArgLysMetAlaPhe 435
963 GTATACCGGTTACGCGCTTACGCGCTTACGCGCTTACGCG 452
435 LTrpAlaValArgPhePheProLeuArgLysMetAlaPhe 452
1013 TCGCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
452 eTrileuValSerGlyLeuValSerGlyLeuValSerGlyLeu 485
1063 AAGGCGGAGTGTATCTCTGCGCGGAGTCCGCTGCTGCTG 1112
469 LysGlySerValSerValThrGlyArgGlnSerProAsnLeuVal 485
1113 TGACGAGCTGTCGATGATGATGATGATGATGATGATGATG 1162
485 GlnAspLysSerSerPheGlnGlnGlnGlnGlnGlnGlnGln 502
1163 CCACCGGCTTACATCATCATCATCATCATCATCATCATCAT 1200
502 LAlaGlyPheIleArgLeuValGlyLeuProMetLys 514

seq_name: swissprot_39:ASSY_BACSU
seq_documentation_block:
ID ASSY_BACSU STANDARD: PRT: 403 AA.
AC 034347:
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
CN Bacillus subtilis.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus; Bacillus; Bacillus.
OC Bacillus; Bacillus; Bacillus.
OC NCB1_TaxID=1423;
OX 111
RN SEQUENCE FROM N.A.
RP MEDLINE=98048467; PubMed=9387221;
RX Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RA "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rnmB-dnaB region."
RT Microbiology 143:3431-3441(1997).
RL -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC -1- PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: PENULTIMATE STEP IN ARGinine BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF008220; AAC00320.1;
CC EMBL: Z99119; CAB14923.1;
CC EMBL: Z99118; CAB14905.1;
CC
CC Subtilisin: BG12570; argG.
CC InterPro: IP001518; Argininosuc_synth.
CC Pfam: PF00764; Argininosuc_synth.1.
CC ProDom: PD003544; Argininosuc_synth.1.1.
CC PROSITE: PS00564; ARGININOSUCCIN_SYN_2; 1.
CC DR PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.
CC Arginine biosynthesis; ligase; ATP-binding; Complete proteome.
KW ARGININOSUCCINATE SYNTHASE; EC:6.3.4.5;
KW SEQUENCE 403 AA; 44817 MW; BFF522A738611F CRC64;

```

800 164 164 164

[illegible]

SWISSPROT\_39:ASSY\_BUCAT

```
seq_documentation_block:
```

STANDARD; PPT. 400

DT 20-AUG-2001 (Bel 40

DT	20-AUG-2001	(Rel. 40, Created)
DT	20-AUG-2001	(Rel. 40, Last

DE	20-AUG-2001 (Rel. 40, ARGENTINO)	Last seq
D1		

...INOSUCCINATE SYNTHASE (EC  
LIGASE).

ARGG OR BU050.

*Buchnera aphidicola* (subsp. *Acyrthosiphon pisum*)

Bacteria: *Proteobacteria*.

NCBI\_TaxID=118099; gamma

SEITE 11

SEQUENCE FROM N.A.  
STRAIN= TOKYO 1000

MEDLINE=20445173; Pubmed 10888888

Shigenobu S., Watanabe H.  
#Co... Hatt...

sequence of the endocytosis

Nature 407:81-86(2000)

-1- CATALYTIC ACTIVITY: ATP + 1

PIROPHOSPHATE + L-ARGININOSU-

PENULTIMATE STEP IN  
SUBUNIT:  
HOMOTETRAMER

-1- SIMILARITY: BELONGS TO THE

----- TO THE A  
This SWEE-DOO-----

between the Swiss-PROT entry is copyright

the European Bioinformatics Institute

use by non-profit institutions

statement is not

or send an email to [licensing@nba.com](mailto:licensing@nba.com).

----- license@sbs-  
EMBL, accession

TABLE; AF001118; BAB12773.1; -.

1

100



DR InterPro: IPR001518; Arginosuc\_synth.  
 DR Pfam: PF00764; Arginosuc\_synth; 1.  
 DR ProDom: PD003544; Arginosuc\_synth; 1.  
 DR PROSITE: PS00564; ARGININOSUCIN\_SYN\_1; 1.  
 DR PROSITE: PS00565; ARGININOSUCIN\_SYN\_2; 1.  
 DR Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.  
 KW SEQUENCE 403 AA; 45072 MW; F2387024C8BDB147 CRC64;

alignment\_scores: Quality: 807.50 Length: 396  
 Ratio: 2.710 Gaps: 5  
 Percent Similarity: 75.253 Percent Identity: 43.687

alignment\_block:  
 US-09-775-693-1 x ASSY\_BUCAI ..

Align seg 1/1 to: ASSY\_BUCAI from: 1 to: 403

```

19 GTGTTCTGCGCTACGTTGGGGCGCTGACACCTCGTGCATCCGCTGTG 68
   |||||||
9  ValValLeuAlaIyrSerGlyGlyLeuAspThrSerAlaIleIleProtr 25
   |||||||
69 GCTGAGGAGACAA...GGCTATGACGTCATTCCTATCTGCGCCACATG 115
   |||||||
25 pleuysgluasnIyrAsnPhelGluValAlaPheValAlaAspIleG 42
   |||||||
116 GCGAG...AAGGAGACTTCGAGAGAGCGACGACGACGACGACGACG 162
   |||||||
42 IyglIserIySylsAspLeuAsnGlyIleGluIySylsSerIleGlu 58
   |||||||
163 GGGGCCAAAAGGTGTCATTCAGATGAGATTCAGACGAGGAGTTGGAGGA 212
   |||||||
59 GlyAlaSerSerCysHisValPheAspLeuIySylsGluIyPheIleGlu 75
   |||||||
213 GTTCATCTGCGCGCGCGCTCAGCGCCAGCGACGATGATGAGACCGTACC 262
   |||||||
75 nIyValIyTrIyProValIleuIySylsThrIyAlaLeuIyrgIySyl 92
   |||||||
263 TCTGCGGACACTCTCTTCCGAGCGCTGATGCGCGCCGACGACGAGAA 312
   |||||||
92 euleuGlyThrAlaMetAlaIarProIleIleAlaIySylsIleValGlu 108
   |||||||
313 ATCCGCCACGCGGAGGGGCGGACGATGTCGCCAGCGCGCCACGAGAA 362
   |||||||
109 LeuAlaLeuAsnIleGlyAlaAsnSerIleuGlyHisGlyAlaThrGly 125
   |||||||
363 GGGGACGATCAGGTCGCGTTTGAAGCTCAGCTCTACTCAGTCGCGCCG 412
   |||||||
125 sGlyAsnAspGlnValArGpHeIuMetAlaIyAlaAlaLeuAlaPro 142
   |||||||
413 AGATAAGAGTCATTGCTCCCGCGGAGATGCTGATTCATCAACCGTTTC 462
   |||||||
142 snLeuAsnValIleAlaIarProIyPtyr....GluTrp....AsnLeu 154
   |||||||
463 AAGGGCGGACATGACCTGATGAGTACGCAAGCAACAGCAGGATCCCAT 512
   |||||||
155 AsnSerArgIySerIleuIySylsIyLeuAspIySylsAsnIleSerTh 171
   |||||||
513 CCGGTCACCTCCCAAGAACCTCGAGCATGAGAGACCTCATGACACA 562
   |||||||
171 rThAlaThrIleuGluIySylsIleIySerIySylsAspIyAsnSerIyPht 188
   |||||||
563 TCGAGTACGAGGCTGATCTCGAGAACCCCAAGACGACGACGCTTCA 612
   |||||||
188 IeSerThrIyGluIyGlyLeuLeuGluAsnProIyPhtAsnIleSerAsnGlu 204
   |||||||
613 GGTCTCTACAGAAAGCCAGAGCCAGCGCAAGCAAGCAAGCCACCTGA 662
   |||||||
205 AspIySerIyPhtIyValAsnProIyGluAspAlaIyProIyGluIyPht 221
   |||||||
663 CATTCGAGATCGATTCAGAAAGAGGCTCCCTGCAAGTGCACACG 712
   |||||||

```

```

221 uTyValSerLeuGlnLeuIySylsGlyCysValIySerValAsnAsnG 238
   |||
713 TCAAGATGGCACCCACACACGACCTCTTGAGCTCTTCATGTACTCG 762
   |||
238 IuIyS.....LeuAsnProIleuIySylsValGlnIuIyGlu 249
   |||
763 AACGAGTCGCGCGGACGACGATGCGCGCGCGCGATTCATGACATGAGAGAA 812
   |||
250 AsnSerIleuGlyAlaIySylsHisGlyIleGlyArGlyIleAspIleIleGluAs 266
   |||
813 CCGCTTCATTGATGAGTCCGAGTATCTACGAGACCCACGACGACA 862
   |||
266 nArgIleuIleGlyLeuIySerArGlyCysTyrgIyIyTrnProGlyGlyT 283
   |||
863 CCATCTTTCACCATGCTCATTCATTCAGATTCGAGGCTTCACGACGACGG 912
   |||
283 hIleIleMetThrAlaIleIySylsAlaIleIyGlnIyGluValIleuIySyl 299
   |||
913 GAGTGGCGAAATTCACACGCGCTGGCTTGAATTCGATGCTGAT 962
   |||
300 GluSerPheArGlyrPArgGluIySylsIleGlyLeuGluMetSerIleIy 316
   |||
963 GTATACGCGTTACGCGCTGAGCTGATTCATTCGCGCCACATGCA 1012
   |||
316 IyIyAspGlyArGlyrPheSerProIleArGlySerIleuIyAlaIy 333
   |||
1013 TCGCCAGATCCACGAGCGAGTGAAGGAGGAAAGTTCAGATGCTGCTC 1062
   |||
333 IAspSerIleuSerLeuGlnIleIyThrGlyGluValIleuIySyl 349
   |||
1063 AAGGCGCAGGTGATCAATCTCGCGCGGAGTCCGACATCTCTCTACAA 1112
   |||
350 IySylsIySerValIyThrAlaValGlnIySylsSerProAsnSerIleuIy 366
   |||
1113 TGAGACGTGCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1162
   |||
366 rGluGluIyAlaIyThrPheGlyGluAspIyValIyIyIySylsIyAsp 383
   |||
1163 CCACGCGTTCATCAACATCAATTCCTCAGGCTGAG 1200
   |||
383 IAspGlyPheIleArGlyLeuPheSerIleuSerIySyls 395
   |||

```

seq\_name: SwissProt\_39:ASSY\_AOUAE

seq\_documentation\_block:

ID	ASSY_AOUAE	STANDARD:	PRT:	401 AA.
AC	067213:			
DT	15-JUL-1999 (Rel. 38, last sequence update)			
DT	20-AUG-2001 (Rel. 40, last annotation update)			
DE	ARGININOSUCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE LIASE).			
GN	ARGG OR AO_1140.			
OS	Aquifex aeolicus.			
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.			
OX	NCBI_TaxID=63363;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VF5.			
RX	MEDLINE=98196666; PubMed=9537320;			
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Sneed G.J., Keller M., Aubay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,			
RA	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus".			
RL	Nature 392:353-358(1998).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +			
CC	PYROPHOSPHATE + L-ARGININOSUCCINATE.			
CC	-1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.			
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			

Quality:	800.50	Length:	404
Ratio:	2.741	Gaps:	3
Percent Similarity:	72.277	Percent Identity:	47.832

```
alignment_block:
  115-00-775 000 1
```

05 05-113-b993-1 x ASSY\_AQUAE  
Align seg 1/1 to: ASSY\_AQUAE

CO: ASSY\_AQUAE from: 1 to: 401

10 AAAGCTCCGTGGTTCTTGCCATACAGTGGCGGCTGGACACCTTCGTCAT 59  
11 |||||  
3 LysLysarValIleLeuValIatrySerGlyIleuAspHisSerIleI 11  
4 |||||  
60 CCTCGTGGCTCAAGCAAGCAAGGCTATAGACGTCAATGCCATTCGCGCA 10  
11 |||||  
19 evalIatrgPrleuInHgIuysgIlyrYgluValIleIrnTrYrInraIa 36  
20 |||||  
110 ACATGGCCGAAGCAAGACACTTCAGGAAGCCAGGAAGACACCTGAG 15  
11 |||||  
36 spValIgluInIgluInIgluInIleuSerIuIleProIuLysIaIargr 52  
160 CTGGGGCCCAAAAAGTGTTCATGAGAGATGAGAGAGGAGATTGTGCA 20  
11 |||||  
53 AlagIyAlaIleGluAlaIleValaIgluAspIeuyrsgIuInrPheAlaI 69  
210 GGAGTTCATCTGCGCGGCGCATCCAGTCCAGCCACTGTATGAGAGCCGT 25  
11 |||||  
69 uAsnIyrcyIeuprOthIrnIeulrYgaIleuAlaIeulrYrIgluIlyst 86  
260 ACCCTCTGGGCACTCTCTTGCCAGGCCCTGTGATGCCCGCAAAACAGTC 30  
11 |||||  
86 yrIProleuThrAlaAlaIleuSerIargrProIeulIleAlaIgluIargIeul 102  
310 GAATTCGGCCAGGGGGGGGGGCAAGATGTGTGTCCAGGCGCCACAG 35  
11 |||||  
103 TyrTrIrnIagIuIySpheAsnIlaAspIlyrAlaIleHsgIySerIhrG 119  
360 AAAGGGGAAGCAAGCAAGTCCGTTTGAAGCTCAGCTGCTACTCACTCGGCC 40  
11 |||||  
119 yIysgIyAsnAspGluValIatrgPheGluIleuSerValITrPalalaIeulAsp 136  
410 CCCAGATAAAGGTCATGTGCTCCCTGGAGAGATGCGCTGAATTCrACACGG 45  
11 |||||  
136 roAspIleGluValIleuAlaIProValrIargIuItrpIuI 148  
460 TTTCAGGGCCGAATGAGACCTATGGAGTACGCCAAAGCAACACGGATTC 50  
11 |||||  
149 PheIySerIargIuInIgluInValaIcIuTrYrAlaIgluIarPheAsnIlePr 165  
510 CATTCCGGTCACTCCCAAGAACCCGTGGAGACATGATGGAGAACCTCATGC 55  
11 |||||  
165 oAlIyysalathIrnIyysgIuIySpIrySerIleAspIargAsnIeuprG 182  
560 ACATCAGCTACGAGCGTGGAAATCTGTGCAAAACCCCAAGACCAACCGCT 60  
11 |||||  
11 |||||  
11 |||||

182 lvalserileglucsyglyproleuclasppttrpgleluupro 198  
610 ccaggtctctacgagagaccgacgagccacgacccacacacccc 659  
199 glnaspralatryglnilethrtnlsrserprogluglnualaaproaspultr 215  
660 tgacattctcgagatcgattgtaaaaaagggctccgtgtaacgtaccca 709  
215 oglutryvalthrvalgilypheglnlucsyglylaserprovaltryleu... 230  
710 acgtcaaggaatggaccaccacccacgacacctcttggactcttcacatgac 759  
231 .....asncllyglutlygtrlyglnuglntrpilsyleuilealaasn 244  
760 ctgacagaaactcgccggcgacagatcgcgtagcccgatgtaacatgctga 809  
245 leuasnlnlilealagllyrghisglyvalgilyarilleanpmevalgl 261  
810 gaaccgcttcatttgatgaatgaatgctccacagatattctacgagaccacgac 859  
261 uasnlatrgleuvalgilyllyllylseratrgglnllytrglnualaprogl 278  
860 gcacacactctttaccatgctcattttagacatcgagagccttcacacatgac 909  
278 lalevalleuvaltryglualatrgasprleuileuileuasp 294  
910 cgg...gaagtgcccaaaatcaaacaaagccctggccttgaatttgcga 956  
295 acgphetrhrphenistyrphelauthrnhsilileprohslutryalaly 311  
957 gctggtgtaacggctttacggccttagccctgagtgtaattgtgtccgcc 1006  
311 sleuvaltrygluileuvaltryphetrhrproleuargglualaleuasp 328  
1007 actgcatcgcccaagctccacgagcagcagtggaaggaagtgacggctgtcc 1056  
328 lapethraalnlylilealgluphealathrlygluvalatrgleu 344  
1057 gtctcgaagggcagggtgtaacattctcgccggcgagctccacatgctctt 1106  
345 leuyltrylsylservaltryvalgllyrghisrserproasnsrle 361  
1107 ctacatgagagactggtgagacatgacgacgagggcgaattatgaccca 1156  
361 utryvalglnuglnleuualathrtryserglusasprlnpneasp 378  
1157 ctgatccacggcggttcacatcaacatcaattccctcagctgaaggaat 1206  
378 leaiacllyshishethrlyllyltryleupoleuylsvalleu 394  
1207 catccgtctccag 1218  
395 glyatrgvalatrg 398

seq\_name: swissprot\_39: ASSY\_LACIA

seq\_documentation\_block:  
ID ASSY\_LACIA STRANDRD: PRT; 398 AA.  
AC P57799;  
DT 20-AUG-2001 (rel. 40, Created)  
DT 20-AUG-2001 (rel. 40, Last sequence update)  
DE 20-AUG-2001 (rel. 40, Last annotation update)  
DE ARKININOSUCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE  
LIGASE).  
GN ARGG.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria: Firmicutes; Bacillus/Clostridium group; streptococcaceae.  
CC Lactococcus.  
RX NBI\_LTaxID=1360;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
XC MEDLINE=21235186; PubMed=11937471.

RA Boletín A., Wincker P., Manger S., Jallón O., Malarme K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*  
 RT *lactis* ssp. *lactis* IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +  
 CC PYROPHOSPHATE + L-ARGININOSUCCINATE.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AE006250; AK04222.1; -  
 CC InterPro: IPR001518; Argininosuc\_synth.  
 CC Pfam: PF00764; Argininosuc\_synth. 1.  
 CC ProDom: PD003544; Argininosuc\_synth. 1.  
 CC PROSITE: PS00564; ARGININOSUCCIN\_SYN\_1; 1.  
 CC PROSITE: PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
 CC Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.  
 KW Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.  
 SQ SEQUENCE 398 AA: 43960 MW: 11410P9C92747712 CRC64.

## alignment\_scores:

Quality: 777.00 Length: 394  
 Ratio: 2.707 Gaps: 4  
 Percent Similarity: 72.843 Percent Identity: 41.117

## alignment\_block:

US-09-775-693-1 x ASSY\_LACLA

Align seg 1/1 to: ASSY\_LACLA from: 1 to: 398

```

1  ATGTCAGCAAGAGCTCGGTGTTCTGCTACAGTGGCGGCTGACAC 50
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1  MetMetGlyAsnLysLysIleValLeuAlaTyrSerLysGlyLeuAspThr 17
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
51  CTCTGTCATCCCTCGTGTGCTGAGCAAGAGCTATGATGCTATGCTT 100
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
17  rSerValAlaValLysTyrLeuThrAspLysGlyPheAspValIleAla 34
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
101  ATCTGCGCAACATGCGCCAGAAAGACTTCGAGAAAGCCAGAGAGAG 150
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
34  spcysMetAspValGlyLysLysAspLeuAsnPheIleHisAspLys 50
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
151  GCACCTAGAGCTTGGGCGCCAAAAGGTTCATTGAGATGTCAGAGAGA 200
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
51  AlaleuGlnValIleValAlaGlnSerValValLeuAspCysLysGln 67
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
201  GTTGTGAGAGAGTTCATCTGCGCGCATTCAGTCCAGCGCAGCTATG 250
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
67  upheAlaLysIlePheValGlyAlaIleLeuLysGlyAsnLeuMetTyr 84
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
251  AGGACCGCTACTCTCTGCGGACCTCTCTGCGAGCGCTGATGCGCGC 300
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
84  LysnLysTyrProLeuValSerAlaLeuSerArgProLeuIleAlaGln 100
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301  AAACAACTGGAATCGCCAGCGAGGGAGGCCAAGTATGTCACCGAG 350
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
101  LysLeuValGlnValAlaLysGlnLysGlyAlaThrAlaIleAlaHis 117
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351  CGCCACAGAAAGGAGACATCAGTCCGCTTGGAGCTCAGCTGCTACT 400
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
117  yCTGThG1yLysGlnAsnAspGlnValAlaArgPheGlnValAlaIle 134
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401  CACTGCGCCCGCCAGATAAGTATGCTCTCCCTGAGAGATGCTGAATTC 450
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

134  erLeuAlaProGlnLeuGlnValIleAlaProValArg.....GlnTyr 148
451  TCAACACGGTTCAGAGCGCCAGCAATGACCTGATGATGACGAAAGACAA 500
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
149  His.....TyrAlaArgLysGlnLysIleGlnTyrAlaAsnGlnAs 162
501  CGGAGATTCCTATCCCGGCTACCTCCCAAGACCCGTGAGCATGTAGTA 550
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
162  nGlnValProIleProAlaAspLeuAspAsnProTyrSerIleAspMet 179
551  ACCTCATGCAATCAGATCTGAGGCTGCAATCTGTAGAGAACCCCAAGAC 600
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
179  snLeuTyrPheArgAlaIleGlnAlaGlyValLeuGlnAsnProTyrPsn 195
601  CAAGCGCTCCAGTCTCTACAGCAAGACCCAGAGACCCAGCCAAAGCC 650
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
196  ThrCysProGlnAspAlaPhePheMetThrAsnSerValGlnAsnAla 212
651  CAACACCCCTGACATTCGTGAGATGCGATGCAAAAAGGGTCCCTGTGA 700
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
212  oAsnGlnAlaGlnPheIleGlnValGlnPheLysGlnLysLeuProIle 229
701  AGGAGACCAACGTCAGATGCGACACCCAGACCTCTGTGAGCTC 750
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
229  lAlaLeu...AsnGlyLysSerLeuGlnLeuHis.....GlnIle 240
751  TTCATGTACCTGAAAGCAAGTCCGCGGCAACATGCGGCTGCTTGA 800
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
241  IleLysLysValAsnIleIleAlaGlyLysIleGlyLysIleGlyLys 257
801  CATGCTGGAACCGCTCATGATGGAATGATGCGGCTGCTGCTTGA 850
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
257  PheIleGlnAsnArgLeuValGlyIleLysSerArgGlnPheTyrGln 274
851  CCCAGACGACCAATCCCTTACATGCTCATTTAGACATGAGGCTTC 900
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
274  yspProAlaAlaIleThrLeuLeuLysAlaHisLysAspLeuGlnAsp 950
901  ACCATGACCGGGAAGTCCGCAAAATCAACAAGGCTGCGCTGAATTT 950
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
291  ThrPheValArgGlnLeuAlaHisPheLysProValLeuGlnAsnGln 307
951  TCTGAGCTGCTGATACCGGCTTACGCGCTAGGCTGATGATGATTTG 1000
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
307  uAlaAsnLeuIleTyrAsnGlyLeuTyrPheAsnProAlaThrLysAla 324
1001  TCCGCGCATGTCGCCCAAGTCCAGAGCGAGAGAGTGAAGGAAGTGCAG 1050
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
324  euIleAlaTyrLeuAspGlnTyrGlnLysValAlaAsnGlyIleValLys 340
1051  GTGTCCGCTCTCAAGGCGCCAGGTACATCTCGGCGGAGTCCCGCACT 1100
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
341  IleLysLeuTyrLysGlyLeuAlaThrProValGlyArgLysSerThr 357
1101  GTCTCTCAATGACAGAGCTGTGTGATGATGATGATGATGATTTG 1150
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
357  nSerLeuTyrSerGlnLysLeuAlaThrTyrThrAlaAspGlnPhe 374
1151  AGCCAACTGATGCCAGCGGCTCATCAATC 1182
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374  spGlnAlaAlaValGlyPheIleLysLeu 384

```

seq\_name: SwissProt\_39:ASSY\_SYN3

seq\_documentation\_block:

ID ASSY\_SYN3 STRAND: PRT: 400 AA.  
 AC P77973;  
 DT 01-NOV-1997 (rel. 35, created)  
 DT 01-NOV-1997 (rel. 35, last sequence update)  
 DT 20-AUG-2001 (rel. 40, last annotation update)  
 DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE-ASPARTATE  
 DE LIGASE).  
 GN ARGG OR SLR0585.

OS *Synechocystis* sp. (strain PCC 6803),  
 CC Bacteria: Cyanobacteria, Chroococcales; *Synechocystis*.  
 OX NOBL\_TaxID-1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97061201; PubMed-8905231;  
 RA Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima T., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA RESeq. 3:109-136(1996).  
 CC -I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARATE = AMP +  
 CC PYROPHOSPHATE + L-ARGININOSUCCINATE  
 CC -I- PATHWAY: PENTULATE STEP IN ARGININE BIOSYNTHESIS.  
 CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D90917; BAA18641.1; -  
 DR InterPro: IPR001518; Arginosuc\_synth.  
 DR Pfam: PF00764; Arginosuc\_synth.1.  
 DR ProDom: PD003544; Arginosuc\_synth.1.  
 DR PROSITE: PS00564; ARGININOSUCCIN\_SYN.1; 1.  
 DR PROSITE: PS00565; ARGININOSUCCIN\_SYN.2; 1.  
 KW Arginine biosynthesis; Liasase; AMP-binding; Complete proteome.  
 SO SEQUENCE 400 AA, 44485 MW, 37BDB758CBB82C05 CRC64;

```
alignment_scores:      Quality: 764.00      Length: 4066  
                      Ratio:    2.709       Gaps:     6  
Percent Similarity:    69.458      Percent Identity: 41.626
```

alignment\_block;

US-09-115-693-1 x ASSY\_SYNY3

Align seg 1/1 to: ASSY\_SYNY3 from: 1 to: 400

```

1  ATGTCACCAAGAGCCTCCGTCGTCCTTCGCTACACATGGCGCGCTCGACAC  50
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1  Metcylalrpalalyslvalvalleualatyrsergylglvalaspth  17
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
51 CTCG...TGCATCCTCGTCGTGGCGTGAAGGACGAT...GACGTCA  94
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
17 rserValCyslIerproIryleumethISgluttrpIglvalIglutIvalI  34
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
95 TTGCCTATCTGCCACACATTTGGCGGCAGAGACATTCGAGAACCCAG  144
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
34 IertrleualalalaspIeuglIglnglIYaspIuleuclIYproIleIln  50
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
145 AAGAGACACATGAAAGCTTGGGGCCAAAGGCTTCATTAGAGATGTGAC  194
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
51 GlulysAlaleuotrGcylglalvalalvalISerleueValIlleasplgIY  67
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
195 CAGGAGATTGTAGAGAGTTCATTCGTGGCGCGCCATTCACATCCAGCGCCAC  244
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
67 sgluglupheValIysglutIryAlalpheuryserIleIlnalalasnAlal  84
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
245 TGTATGAGACCGCTACCTCTCGGCGACCTCTTCGACAGGCCCTGATC  294
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
84 eutIryglusnasnargIYrProIeusertrAlaleualalalargProIeulIle  100

```

295 GCGCGCAACAGAGGGAATCGGCCGACGGGAGGGGCGCAATGATGTGC 344  
101 AAtyStetueuValGluAlaAlaGluTyTyrGluAlaAspAlaValAl 117  
345 CCACGGGCGCCACAGGAAGGGAGGACATGAGGTCCGGTTGAGCGACGT 394  
117 aNIGlIcYugThtGluYluelYAsnAspInValArgHephSrlseST 134  
395 GGTACGTACGTGGGCCCCGAGTAAAGTCATTGTCTCC...TGG 435  
151 Lysmet.....SerArgGluGutHrIleAl 159  
486 GTACGCAAGACAAACAGGGATTCCATCCGGTCACTCCAGAACCCGT 535  
159 aTyrGluYgluAlaTyrGluYgluValGluSerProValLysLysSerProI 176  
536 GAGGACATGATGAGAACCCTGATGCATACATCACTCAAGGCTGGAAATCTG 585  
176 yTerIleAspArgHsnIleleuGluArgSerIleGluAlaGlyProLeu 192  
586 GAAACCCCAACAAACCAAGCGCTCCAGGTCCAGTCCAGAACGACGAGA 635  
193 GluAspProMetThrGluProThrGluGluIleTyrLeuMetThrIysAl 209  
636 CCGACGCAAGACCCCGCCACACCCCGACATTTCTCGAATCGAGTTAA 685  
209 AlleAlaAspHrProAspArguProGluTyrValAspIleGlyPheGlu 226  
686 AAGGGTCCCTGTGAAGGTGCACAGCTCAAG...GATGGACCAACCCAC 732  
226 yGluIleProValSerIleuAsnGluYalIleMetuAspProValHr... 241  
733 CAGACCTCTTGAGCTCTTCATGTACTGTGAACGAAGTCCGGGCAACA 782  
242 .....LeuValGluAlaGlyLeuAsnGluIleAlaGlyAsnH 253  
783 TGGCGTGGCGGATTGACATCGGAGGAACCCCTTGAGAAATGAGT 832  
253 scIyAlaGluTyrLeuAspMetValGluAsnArgValAlaGlyIleLys 270  
833 CCGGAGTATCTAGACACCCGACGACGACATCTCTTACCATGCTCAT 882  
270 eArgGluIleTyrGluAlaProIaLeuIleuValIleAlaHIs 286  
883 TTAGACATCGAGGCTTCAACATGACCGCGAAGGCGCAAAATCAACA 932  
287 ArgAspLeuGluSerIleuHrGluInHrAlaAspValInHrIstYrLys 303  
933 AGGCGTGGGCTTAAATTTCGTAGCTGGTGTATACCGGTTTACGGCTA 982  
303 nThValGluGluIleTyrSerGluIleuIleTyrArgGlyLeuHrPyr 320  
983 GCGCTGAGTGAATTAATTTGCCGCCCACTGCATGCCAATGCCAGAGCGCA 1032  
320 eProLeuYsgluAlaLeuAspAlaPheIleValYsthGlnGluArg 336  
1033 GTGGAAGGAAACTGACAGGTGCCGTCCCTCTCAAGGCGCAGGTGATCATCC 1082  
337 ValThrGlyMetValArgValLysPheHephYsgLysnAlaSnValAl 353  
1083 CGCGCGGAGTCCACACGTCTCTCTACATGAGGACAGCTGTGAGCATGA 1132  
353 aGlyAlaGlySerAspTyrSerIleTyrAspAlaGluLeuAlaHrTyrG 370  
1133 ACGTGCAGGAGTATATATAGCAACTGATGCACCGGGTTCATCAACATC 1182  
370 TyMetGluAspInHephAspHisLysAlaAlaGluGlyPheIleTyrIle 386  
1183 AATTCGCTCAGGCTGACG 1200

387 TrpGlyLeuProThrLys 392

seq\_name: swissprot\_39:ASSY\_ARCFU

seq\_documentation\_block: ASSY\_ARCFU STANDARD; PRT; 390 AA.

AC 028032: 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE  
 DE LIGASE).  
 CN ARGG OR AF2252.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 CC NCBI\_TaxID=2234;  
 OX 11  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RC MEDLINE=98049343; PubMed=9389475;  
 RX Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 Klenk H.-P., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,  
 Richardson K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,  
 Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 Fierman S., Ketch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 Peterson S., Ketch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 Overbeek R., Gocayne T., Artach P., Kaine B.P., Sykes S.M.,  
 Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 Venter J.C.;  
 RA "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 380:364-370(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +  
 CC PYROHOSPHATE + L-ARGININOSUCCINATE.  
 CC -1- PATHWAY: PENULTIMATE STEP IN ARGININE BIOSYNTHESIS.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AE000949; AAB89005.1;  
 DR TRNG: AF2252;  
 DR InterPro: IPR001518; Argininosuc\_synth.  
 DR Pfam: PF00764; Argininosuc\_synth. 1;  
 DR ProDom: PD003544; Argininosuc\_synth. 1;  
 DR PROSITE: PS00564; ARGININOSUCCIN\_SYN\_1; 1;  
 DR PROSITE: PS00565; ARGININOSUCCIN\_SYN\_2; 1;  
 DR Arginine biosynthesis; ligase; ATP-binding; Complete proteome.  
 KW Arginine biosynthesis; ligase; ATP-binding; Complete proteome.  
 SEQUENCE 390 AA: 44104 MW: 313408469696EBD CRC64;

alignment\_scores: Quality: 750.50 Length: 407  
 Ratio: 2.544 Gaps: 12  
 Percent Similarity: 72.482 Percent Identity: 43.735

alignment\_block: US-09-775-693-1 x ASSY\_ARCFU

Align seg 1/1 to: ASSY\_ARCFU from: 1 to: 390

19 GTGGTCTGGCTACAGTGGCGCTGACACCTGTCATCTCGTGTG 68  
 |||||||

3 ValValLeuSerTyrSerGlyLeuAspThrThrValCysIleProLe 19  
 69 GCTGAAGAGAACAA...GGCTATGAC...GTCAATGGCTATGTGGCCACAA 112  
 |||||||  
 19 uLeuysGluLysTyrGlyPheAspGluValIleThrValThrValAspI 36  
 113 TTGGCCACAGAGAA...GACTTCAGAGAACCCGAGAACAGGCACTGTAG 159  
 |||||||  
 36 LeGlyLysProGluAlaAspIleLysGlnAlaGluLysGlyLysIys 52  
 160 CTGGGGCCCAAAAGAGTTCATTCAGAGATTCAGCGAGGAGTTTGTGA 209  
 ||| |||  
 53 Tyr...AlaAspLysHisTyrThrIleAspAlaLysGluPheValas 68  
 210 GGAGTTCATTCAGCGCGCCGCTCAGTCCAGCGCACCTGTATGAGACCGCT 259  
 : : : : :  
 68 Pser...LeuPheMetLeuLysAlaAsnGlyAsnTyrIleLys...T 83  
 260 ACCTCCTGGGACCTCTCTTCGCCAGCGCTGCATCGCCGCAACAGATG 309  
 |||||||  
 83 YrValLeuGlyThrAlaLeuAlaArgProLeuIleAlaGluLysVal 99  
 310 GAATCGCCCGCCGAGGAGGCGCAAGTATGTTCACCGCGGCGCACAG 359  
 |||||||  
 100 GluValAlaLysLysGluGluAlaGluAlaValAlaHisGlyCysThrG 116  
 360 AAGAGGAGCAGATCAGTCCGCTTTCAGCTCAGCTCTACTGCTGCCC 409  
 |||||||  
 116 LysGlyLysAspGlnLeuArgPheGlu...AsnIlePheA 129  
 410 CCGAG...ATAAGGTCATGTCTCCCTGAGAGATGCTGATTCAT 453  
 |||  
 129 rGlnHisGlyPheLysValIleAlaProValArgGluLeuAsnLeuThr 145  
 454 ACCCGTTCAGAGCGCCGCAATGATGAGTTCAGTTCAGTTCAGTTCAG 503  
 : : : : :  
 146 ArgLysIleProLeu...IleGluLysAlaArgGlnHisG 157  
 504 GATTCCTCCGCTGATCCTCCAGAACCGCGTGGAGTATGAGAAC 553  
 ||| : : : : :  
 157 YLleGluValProAlaThrLysGluLysProLysIleAspGluAsn 174  
 554 TCATGACATCAGTCAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 603  
 ||| |||  
 174 eutPThrArgSerValGluGlyLysLeuLysProSerPheGlu 190  
 604 GCGCTCCAGCTCTACACAGAGACCCAGCCAGCCAGCCAGCCAGCC 653  
 ||| : : : : :  
 191 ProProGluAspIleTyrGluTyrPheAlaSerProGluLysAlaPro 207  
 654 CACCCGTCATTCGAGATCGATGATCAAAAAGGAGTCCCTGAGAG 703  
 : : : : :  
 207 PylsProGluIleValLysIleAspPheGluLysGlyValProValAla 224  
 704 TGACCAACGTCAAGATGGCAGCCAGCCAGCCAGCCAGCCAGCCAG 753  
 : : : : :  
 224 euaAsnAspGluArgMetGly...GlyPheGluLeuIle 235  
 754 ATGACTTCAGAGCAAGTCCGCGGAGCAAGATGGCGGCGCTATTCAT 803  
 |||||||  
 236 LysAlaLeuAsnGluLysGlyLysHisGlyValGlyArgThrAspMe 252  
 804 CTTGAGAGACCGCTTCATTCGAGTTCAGTTCAGTTCAGTTCAGTTCAG 853  
 : : : : :  
 252 tLleGluAspArgValLeuGlyLeuLysAlaArgGluAsnTyrGluHisP 269  
 854 CAGCAGCAGCAGCAGTTCATTCATTCATTCATTCATTCATTCATTCAT 903  
 |||||||  
 269 roAlaAlaThrIleLeuLysThrAlaHisArgAspLeuLysLeuVal 285  
 904 ATGACCGGAGAGTCCGCAAAATCAACAGAGCGCTGAGCAATTCGC 953  
 : : : : :  
 286 LeuSerArgArgGluLeuLysPheLysLysPheValGluLysIlePro 302

```

954 TGACGTGCTATACCGGTTTACGCGCTAGCCCTGAGTGAATTTGCC 1003
302 agluLeuValTyrGlyLeuValAsnAspProLeuPheAspAlaLeu 319
1004 GCCATGTCATCGCCAGTCCCGAGGAGGAGTGAAGGAAAGTCCAGCTG 1053
319 snAlaPheLeuAspLysThrGlnGluArgValThrGlyTyrPValLysVal 335
1054 TCCGTCCTCAAGGCGCAGGTGATCATCTCGCGCGAGGAGTCCCTGAG 1103
336 LysLeuTyrLysGlySerAlaValAlaValAlaThrAsnSerProTyrAl 352
1104 TCTTCAACATGAGAGCTGTCGACATGACAGTGCAGGTGATTTAGAGC 1153
352 aLeuTyrSerGlnGluLeuValSerPhe.....AspThrGlnLys 365
1154 CAACGTGAT.....GCCACGCGGTTCATCATCAATCATCTCCCTCAGG 1194
365 erLLeAspGlnArgLeuAlaGluGlyPheAlaAlaPheHisGlyLeuGln 381
1195 CTGAGAGAAATATCATGCTCTC 1215
382 GlyArgLeuPheArgArgLeu 388

```

seq\_name: SwissProt\_39:ASSY\_METUA

seq\_documentation\_block:

```

ID ASSY_METUA STANDARD; PRT; 395 AA.
AC 060174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-NOV-2001 (Rel. 40, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
GN ARG OR MJ0429
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
OX Methanococcus.
RN NCBI_TaxId=2190;
RC SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=8686087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Scott J.L., Geoghegan N.S.M., Weinstock K.G., Merrick J.M., Glodek A.,
RA Utterback L.R., Kelley J.M., Peterson J.D., Fuhmann J.L., Nguyen D.,
RA Cotton M.D., Roberts C.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE =
CC -1- PATHWAY: PENTULTIMATE STEP IN ARGININE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: U67494; AAB8414.1.
DR TIGR: MJ0429;
DR InterPro: IPR001518; Arginosuc_synth.
DR Pfam: PF00764; Arginosuc_synth.1.

```

```

DR ProDom: PD003544; Arginosuc_synth.1.
DR PROSITE: PS00564; ARGININOSUCCIN_SYN.1; 1.
DR PROSITE: PS00565; ARGININOSUCCIN_SYN.2; 1.
KW Arginine biosynthesis; Ligase; ATP-binding; Complete proteome.
SQ SEQUENCE 395 AA; 44723 MW; 655DA7AC06E7A7F CRC64;

```

# alignment\_scores:

Quality:	Ratio:	Length:
Percent Similarity: 70.647	2.630	402
Percent Identity: 43.035	Gaps: 9	

## alignment\_block:

US-09-775-693-1 x ASSY\_METUA

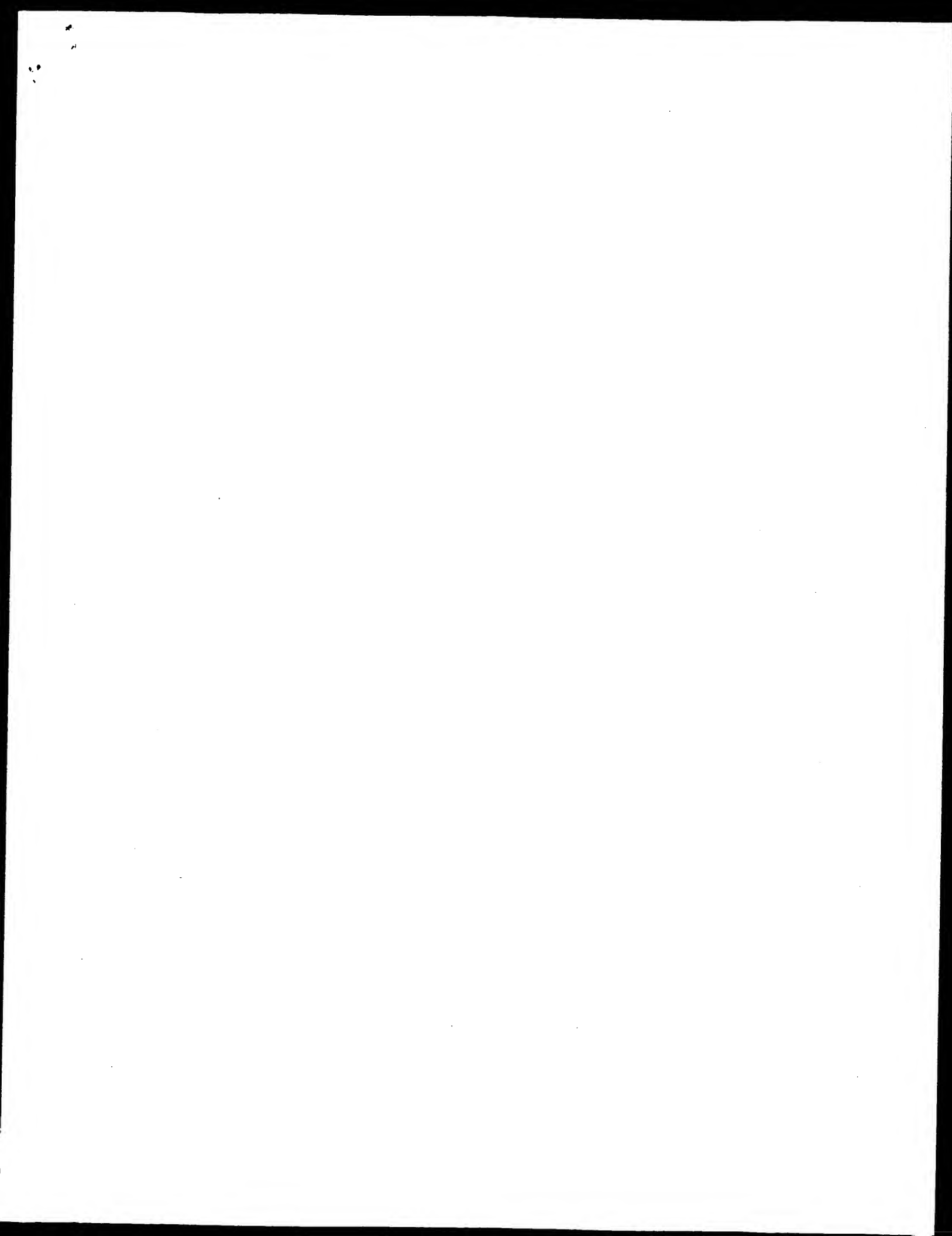
Align seq 1/1 to: ASSY\_METUA from: 1 to: 395

```

22 GTTCGTGCGCTACAGCTGGCGCTGACACCTGTCGTCGTCGTCGCT 71
6 ValLeuAlaTyrSerGlyLeuAspThrSerCysLeuLysLeu 11
72 GAAGGACAA...GGCTATGACGTTCATTCGCTATTCGCCAATGGCC 118
22 uGlnAspLysTyrGlyTyrLysValValSerValCysValAspValGly 39
119 AG...AAGGAGACTTCGTCGAGAGAGCCAGAGAGAGAGAGAGAGTGG 165
39 InProGlnGluGluLeuLysGluValGluGluLysAlaLysLysLeu 55
166 GCCAAAAGGTTCATTCGATTCGATTCGATTCGATTCGATTCGATTCG 215
56 ValLeuLysHisTyrThrLeuAspAlaLysGlnGluLysPheValLys 72
216 CATCTGCGCGCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 265
72 rLLePheArgAlaLeuLysAlaAsnAlaMetLysGly...TyrProL 88
266 TGGGCACTCTCTCTCCAGCGCTGATCGCGCGCGCGCGCGCGCGCGCG 315
88 euserThrAlaLeuAlaArgProLeuLeuAlaHisLysValValGlu 104
316 GCCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 365
105 AlaGluGluValGlyAlaGluAlaValAlaHisGlyCysThrGlyLys 121
366 GAACATGAGTCCGCTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 415
121 YasnAspLysPheArgPheGluThrThrLeuArgLysAlaProHisL 138
416 TAAAGTCATTCGCTCCCTGAGAGATGCTGATTCACACCGGTTGAG 465
138 euLysLysLeuAlaProLeuArg.....AspLeuAsnLeu 149
466 GCCCGCAATGAGCTGATGAGTACGCAAGACAGGAGTCCCATCCATCC 515
150 ThrArgAlaGluGluLeuLeuTyrAlaLysGluLysLysLysLysL 166
516 GGCATCTCCCAAGAACCCGTGACATGATGAGAACCTCATGACATCA 565
166 o...ThrLysSerLysTyrSerLysLeuAsnLeuTyrGlyArgS 182
566 GCTACGAGGCTGAGTTCGAGAACCCCAAGAACCCCAAGAGCTCCAG 615
182 erLLeGluGlySerGlnGluAsnProAspPheValProProGluGlu 198
616 CTCTACAGAAAGCCAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 665
199 IleTyrAlaTyrPheLysAsnProValGlu...AspLysGluGluGlu 214
666 TCTGAGATGAGTTCACAAAAGGAGGCTCCGTGAGAGTGCACCACTCA 715
214 eValGluLysGluPheLysGluGlyValProValAlaLysGlnGlu 231

```

```
716 AGATGGACACCCAGACCTCTTGAGCTCTTCATGTACCTGAC 765
||
||
231 ys.....LeuGluProValGluLeuIleLysLysAlaasn 242
||
766 GAAGTCGGCGGCAAGCATGGCGCTGATGACATGTGGAGAACCG 815
|||||
243 GluIleAlaGluLysHisGluValGluArgIleAspIleIleGluAspAr 259
||
816 CTTCATTTGATGAAGTCCCGAGGTATCTACGAGACCCAGCAGCACCA 865
||
259 GluIleGluLeuLysSerArgGluAsnTyrGluCysProGluAlaValL 276
||
866 TCCTTTACCATGCTCATTTAGACATCGAGCGCTTCACCATGACCGGAA 915
||
276 euleuLeuThrAlaHisLysAlaLeuGluGluValLeuThrArgAsp 292
||
916 CTCGCGAAATCAACACAGGCTGCGGCTTGAAATTTGCTGAGCTGCTGA 965
||
293 GluLeuArgPheLysGluIleValAspSerLeuTyrGluLeuIleTyr 309
||
966 TACCGGTTTACGGCCTAGCCCTGAGTGAATTTGTCCGCCACTGCATCG 1015
||
309 rLysGlyLeuTrpPheAspProLeuArgGluAspLeuAspAlaPheIleA 326
||
1016 CCAAGTCCCGAGGACCGAGTGGAGGAAGTGCAGGTGCCGTCCTCAAG 1065
||
326 sPluStrIngInuArgValThrGlyThrValLysValLysLeuPheGly 342
||
1066 GGCAGGTGTACATCTCGGCGCGGAGTCCCACTGTCTCTTACAAATGA 1115
||
343 GlyThrAlaArgValValGlyArgAspSerProTyrAlaLeuTyrSerLy 359
||
1116 GGAGCTGTGAGCATGTGACGTGCAGGCGTATGAGCCAACTGATGCCA 1165
||
359 sGluLeuValSerPheasp...GluLysGluIleAspGluLysGluLeuA 375
||
1166 CCGGTTTCATACATCATTCCTCTCAGCGTGAAGAAATATCATTCGTCTC 1215
||
375 IaGlyMetVal.....LysTyrHisGlyLeu 383
||
1216 CAGAGC 1221
||
384 GlnAla 385
```





```

00-775-603-1 to: SPTREMBL_17:* out_format : p

```

About: Results were produced by  
Copyright (c) 1993-2000 CompuGen Ltd.

```

-DB_CGNUM2_1/VSPFO_spool/050571      SUFFIX=rspt      -GAPOP=12.000
-DB_SERVERMEM.17      -OFMT=fastan      -LOOPT=0.000      -LOOEEXT=0.000
-OGAEXT=4.000      -MINMATCH=0.100      -XGAPOP=10.000      -GGAEXT=0.500
-OGAOP=4.500      -OGAEXT=0.050      -YGAPOP=10.000      -GGAEXT=0.500
-EGAPOP=6.000      -EGAPEXT=7.000      -SRPT=1      -MARRIX=blotsum62
-DELPO=6.000      -DELEXT=7.000      -FOCALIGN=200      -THR_SCORE=pcr
-TRN_humand40.cdi      -list=45
-TRN_MAX=100      -TRN_MIN=0      -ALIGN=15      -MODE=LOCAL      -OUTFM=ps
-NOH=ext      -MINLEN=0      -MAXLEN=2000000000
-USER=US09375693.cgcnl_1_146      -NCPU=6      -ICPU=3      -LONGLOG      -NO_XLPPX
-WAIT      -THREADS=1

```

```
Search information block:
Query: US-09-775-693-1
Query length: 1239
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 125.310000
```

Sequence	Strid	Orig	Zscore	EScore	Len	Documentation
SP_bacteria:09X2N1	+	1226.50	1443.71	9.4e-73	409	09X2N1 thermotoga maritima .arg
SP_bacteria:09R4U1	+	1034.00	1216.81	4.2e-60	402	09R4U1 deinococcus radiodurans .arg
SP_bacteria:09R4U1	+	948.50	1115.84	1.7e-54	408	09R4U1 caulobacter crescentus .arg
SP_bacteria:09H4Y4	+	927.50	1091.13	4.1e-53	405	09H4Y4 pseudomonas aeruginosa .arg
SP_bacteria:09PHK7	+	892.00	1049.24	8.9e-51	406	09PHK7 campylobacter jejuni .arg
SP_bacteria:09K8Z0	+	865.00	1017.33	5.3e-49	409	09K8Z0 bacillus halodurans .arg
SP_bacteria:09K4Z3	+	818.00	962.00	6.5e-46	404	09K4Z3 moraxella sp. 2674 .arg
SP_plant:09S2X3	+	815.00	926.46	1.0e-45	498	09S2X3 arabidopsis thaliana .arg
SP_bacteria:09K4Y8	+	776.00	912.46	3.7e-43	401	09K4Y8 moraxella sp. 2693 .arg
SP_bacteria:09YCV7	+	761.00	894.83	3.6e-42	404	09YCV7 staphylococcus aureus .arg
SP_bacteria:09KMT8	+	751.50	883.56	1.5e-41	401	09KMT8 vibrio cholerae .arg
SP_bacteria:09CC10	+	785.50	805.82	3.3e-37	399	09CC10 mycobacterium leprae .arg
SP_archaea:09HX31	+	656.50	771.47	6.2e-32	406	09HX31 sulfolobus solfataricus .arg
SP_human:04K348	+	604.50	717.81	1.3e-27	396	04K348 homo sapiens (human) .arg
SP_archaea:09NM02	+	539.50	653.68	7.3e-23	448	09NM02 sus scrofa (pig) .arg
SP_mammal:02R262	+	466.00	559.14	1.3e-19	447	02R262 neisseria meningitidis .arg
SP_bacteria:09SU01	+	417.50	488.73	2.1e-19	441	09SU01 neisseria meningitidis .arg
SP_bacteria:09J0M1	+	414.50	445.43	1.2e-16	487	09J0M1 streptomyces coelicolor .arg
SP_bacteria:09PC47	+	372.00	434.43	1.3e-16	401	09PC47 xyloella fastidiosa .arg
SP_bacteria:09JEM9	+	367.50	430.70	9.7e-12	167	09JEM9 methanobacterium thermophilum .arg
SP_archaea:09V223	+	296.50	356.94	9.7e-09	177	09V223 carboxydothermus hydrog
SP_bacteria:09HNB9	+	251.50	300.99	0.0004	242	09HNB9 erwinia chrysanthemi .arg
SP_bacteria:09G8M0	+	181.50	215.71	0.0034	608	09G8M0 herpessivus papio .arg
SP_virus:09O5K9	+	167.50	191.77	0.0037	580	09O5K9 homo sapiens (human) .arg
SP_human:09S0M8	+	159.50	180.36	0.0126	713	09S0M8 triticum aestivum (wheat)
SP_plant:09S0M2	+	157.00	171.95	0.0198	1336	09S0M2 homo sapiens (human) .arg
SP_human:09C0J8	+	156.50	179.74	0.0191	510	09C0J8 arabidopsis thaliana (mc
SP_plant:09C0J8	+	155.50	175.02	0.0233	766	09C0J8 triticum aestivum (wheat)
SP_plant:09S0D5	+	153.00	179.69	0.0307	319	09S0D5 homo sapiens (human) .arg
SP_human:09UF55	+	153.00	167.30	0.0362	1325	09UF55 leishmania major .arg
SP_liver:09E4R6	+	149.00	194.45	0.0434	34	09E4R6 fugu rubripes (japanese f
SP_vertebrate:05T591	+	148.50	174.72	0.0604	307	05T591 homo sapiens (human) .arg
SP_human:09R705	+	148.50	172.76	0.0871	274	09R705 zea mays (maize) .arg
SP_plant:09R715	+	142.00	166.71	0.1623	319	09R715 deinococcus radiodurans .arg
SP_bacteria:09R4V01	+	142.00	160.16	0.1771	678	09R4V01 homo sapiens (human) .arg
SP_human:09A850	+	140.50	153.43	0.2311	1190	09A850 homo sapiens (human) .arg
SP_human:09S621	+	140.00	165.77	0.2135	271	09S621 mus musculus (mouse) .arg
SP_rodent:09CCT8	+	140.00	159.75	0.2355	542	09CCT8 myxococcus xanthus .arg
SP_bacteria:06M872	+	140.00	152.96	0.2555	1182	06M872 homo sapiens (human) .arg
SP_human:09A945	+	140.00	152.96	0.2555	1182	09A945 homo sapiens (human) .arg

```
seq-documentation_block: 409 AA
cov2a1 PRELIMINARY; PRT;
```

AD	Q9X2A1: 1999 (TREMblrel. 12. Created)
DT	01-NOV-1999 (TREMblrel. 12. Last sequence update)
DT	01-NOV-1999 (TREMblrel. 12. Last annotation update)
DT	01-JUN-2001 (TREMblrel. 17. Last annotation update)
DE	ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE LIGASE).
GN	TML780.
OS	Thermotoga maritima.
OS	Bacteria; Thermotogales; Thermotoga.
OX	NCBI_TaxID=2336;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRFAM-MSB8 / DSM 3109;
FX	MEDLINE=99287316; PubMed=10360571;
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA	McDonald L., Urbach T.R., Matk J.A., Linher K.D., Garrett M.M.,
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA	Heidelberg J., Sutton G.G., Fleischman R.D., Eisen J.A., White O.,
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from
RT	genome sequence of Thermotoga maritima.";
RL	Nature 399:323-329(1999).
RL	-1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC	pyrophosphate + L-ARGININOSUCCINATE.
CC	-1- PATHWAY: UREA CYCLE, PENITILLATE STEP OF THE ARGININE BIOSYNTHETIC
CC	PATHWAY.
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC	-1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
CC	EMBL: AE001816; AAD36844.1; -.
DR	TIGR: TML780; -.
DR	InterPro: IPR001518; Argininosuc-synth.
DR	Pfam: PF00764; Argininosuc-synth. 1.
DR	Pfam: PD003544; Argininosuc-synth. 1.
DR	PROSITE: PS00564; ARGININOSUCCIN-SYN_2: 1.
DR	PROSITE: PS00565; ARGININOSUCCIN-SYN_2: 1.
KW	ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW	urea cycle.
SQ	SEQUENCE 409 AA: 46054 MW: ECDC08575E962482 CRC64;

```

alignment_scores:      Length: 400
                       Quality: 1
                       Ratio: Gaps:
Percent Similarity: 3.683    Percent Identity: 59.012
                       82.222

```

```
alignment_block: ..
               09X2A1
```

US-09-7/5-693-1 A X  
Align seg 1/1 to: Q9X2A1 from: 1 to: 409  
CTGGCGGCGCTGGACACCTCGTCAT 59

[illegible]

[illegible]

alignment_scores:	
Quality: 1034.00	Length: 411
Ratio: 3.272	Gaps: 6
Percent Similarity: 76.886	Percent Identity: 50.122
alignment_block:	

```

US-09-775-693-1 x Q9RMJ4 ..
Align seg 1/1 to: Q9RMJ4 from: 1 to: 402
7 ACCAAGGCTCCGATGGTCTGACCTACAGTGGCGGCTGAGACACTGCTG 56
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

2 SerLysGluValIleValLeuAlaTyrSerIleGlyLeuAspThrSer11 18
57 CATCTCGTGTGCTGAAG...GAACAAGCTATGACGTCAATTCCTATC 103
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
18 eIleLeuLysTrpLeuGlnThrGluArgAsnIleValAlaCysPheT 35
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
104 TGCCCAACATTTGGCCAGAGAAAGACTTCGAGAGCCAGGAGAGAGCA 153
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
35 hLlaAspLeuGlnIleGluValAlaGluAlaIleValAlaValLysAla 51
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
154 CTGACCTGGGGGCCAAAGGTTCATTTGAGATGTCAGCAGGAGGAT 203
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
52 LeuAsnThrGluAlaValAlaIleValAlaLeuAspLeuArgGluLuph 68
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
204 TGTGAGAGATTCATTCGGCCGCCATCCAGTCCAGCCGACCTGTATGAG 253
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
68 eValArgAspTyrValIlePheProMetMetArgSerSerAlaLeuTyrGluG 85
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
254 ACCGCTAACCTCCGGGACCTCTCTTCGACAGGCCCTCATCGCCGCAAA 303
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
85 LyrTyrTyrLeuLeuGlnIleThrSerIleAlaArgProLeuIleAlaLysLys 101
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
102 MetValGluIleAlaGluLysGluValAlaValAlaIleSerHisGlyAl 118
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
354 CACAGAGAAAGGGAGACATCATGCTGGGTTTGAGCTCAGCTACTGAC 403
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
118 aThrGlyLysGlyAsnAspGlnValArgPheGluMetSerAlaTyrAla 135
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
404 TGGCCCCCAGATPAAGATTCATTCCTCCGAGATTCCTGATTCAC 453
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
135 eulysProAspIleValIleThrValAlaLeuProIleArgAspIlePasp 149
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
454 AACCGTTCAAGGGCGCAATGACCTGATGAGTACGCAAGCAACAGCG 503
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
150 .....PheGlnIleArgAlaAspLeuGluAlaPheAlaArgGluHisGly 164
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
504 CATTCCTCCATCCCGCTCCACATCCGAGACCGGTGAGATGAGAGAAC 553
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
164 YIleProValProThrThrLysLysAspProIlePheSerMetAlaAsnM 181
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
554 TCATCATCATCATGCTACGAGGCTGGAATTCCTGGAACCCCAAGAACAA 603
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
181 eIleHisIleSerTyrGlnIleGlyLysProLeuGluAspProIlePheGlu 197
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
604 GCGCTCCAGGCTCTACACAGCAAGACCCAGCAGCAAGCCCAAGCCCA 653
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
198 ProIleThrHisMetPheLysLeuThrValAlaAsnProGluAspAlaPro 214
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
654 CACCCCTGATTCGAGATCGAGTCAAAAGAGGGTCCCTGCTGAAG 703
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
214 tGluAlaGluTyrValGluIleGluTyrValAlaAsnIleAspProValSer 231
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
704 TGACCAACGTCAGAGTGGACACCCAGACCTCTTGAGACTCTC 753
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
231 leAsnGly.....GluGlnLeuSerProAlaAlaLeuLeu 242
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
754 ATGTACTCTAAGAGTGGCGGCAAGATGCGGCGGCTATTTGACAT 803
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
259 uValGluAsnArgPheValGlyMetLysSerArgGlyValIleTyrGluThr 276
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
854 CAGCAGGACACATCTTACATGCTCATTTAGACATGAGGCTTAC 903
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
276 roGlyGlyThrLeuLeuTyrHisAlaArgAlaValGluSerIleThr 292
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
904 ATGACCGCGGAAGTGGCAAAATCAACAGAGCTGGGCTGAAATTCG 953
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
293 LeuAspArgGluValLeuHisGlnArgAspAlaLeuGlyProLysTyrAl 309
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

954 TGAGCTGTGTATACCGGTTATACGGCTACGCCCTGAGTGTGA..... 996
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
309 aGluLeuValIleTyrAsnGlyPheThrPheAlaProGluArgGluAlaLeuG 326
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
997 .....TTTCTCCGACATCAGCCAGTCCAGGAGCAGGAGGAGG 1041
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
326 InValIlePheAspHis...ValAlaLysSer.....ValThrGly 338
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1042 AAAGTCAGCTGTCCGTCCTCAAGGCCAGGTGTACATCTCGGCCGGA 1091
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
339 ThrAlaArgLeuLeuLeuTyrLysGlyAsnLysIleValAlaGlyArgLys 355
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1092 GTCCGACATCTCTCTCTACATGAGAGAGCTGTGAGCATGAACTGACG 1141
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
355 sAlaGluArgSerLeuTyrAspLysAspLeuValSerPheGluAlaGlyG 372
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1142 GTGATTTATGAGCCCACTGATGCCACCGGTTTCATCAACATTCCTCT 1191
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
372 LysAspTyrAsnGlnHisAspAlaGlyAlaPheIleLysLeuAsnSerLeu 388
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1192 AGCCTGAGGAATATCATCTCTCCAGAGCAAG 1224
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
389 ArgMetArgValGlnLysArgValGluAspLys 399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

seq\_name: sp\_bacteria:Q9ABU1

seq\_documentation\_block:

ID Q9ABU1 PRELIMINARY; PRT; 408 AA.

AC Q9ABU1;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE ARGININOSUCCINATE SYNTHASE.

GN CC0129.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_TaxID=69394;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocky I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uiterback T., Tran K., Wolf A., Yamthuevan J., Ermoiaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RA "Complete genome sequence of Caulobacter crescentus."

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

RK EMBL: AB005687; MAK22116.1; ..

DR TIGR: CC0129; ..

KW Complete proteome.

SQ SEQUENCE 408 AA; 45330 MW; 06571CHDC38B0FBD CRC64;

alignment\_scores:

Quality: 948.50 Length: 401

Ratio: 3.090 Gaps: 8

Percent Similarity: 76.559 Percent Identity: 50.125

alignment\_block:

US-09-775-693-1 x Q9ABU1 ..

Align seg 1/1 to: Q9ABU1 from: 1 to: 408

```

19 GTGGTTCTGCGCTACAGTGGCGCTGAGACCTCTGTCATCTCTGCTG 68
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
9 ValValLeuAlaTyrSerIleGlyLeuAspThrSerIleIleLeuLysTr 25
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
69 GCTGAGAG...GAACAAGCTATGACGTCAATTCCTGAGCAACATTCG 115
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

1004 GCCAAGCAGTCGCCAGCAGCTCCCAAGACGACGCTGAAAGCAAACTGACGGTG 1053  
 333   :::|||||   |||||:::|||||   |||||:::|||||   |||||:::|||||  
 333   lnalaatalleaptryrsegltnasplvysvalthgIyArValrYval 349  
 1054 TCCGTCCTCAAGGCCAGGCTGTACATCCCTCGGCGGGAGTCCCACTGCTC 1103  
 350   :::|||||:::|||||   :::|||||:::|||||   :::|||||:::|||||   ||  
 350   LyLeuLyLyLyseLylnasvAlnhrvalIleglYalrgluseprYrse 366  
 1104 TCCTCATATGAGGAGCTGTGTAGCATGAACTGCAGGGTGAT.....T 1147  
 366   |||||:::|||||   :::|||||:::|||||   :::|||||   :::||||  
 366   rLeuYrAspGlnAspLeValThrPheGlu...GluSlyLySvalAlat 382  
 1148 ATGAGCGCACTGATCCACCGCGTTCATCAACATCAATTCCTCGACGCTG 1197  
 382   |||||   |||||   |||||:::|||||   :::|||||:::|||||   |||||  
 382   YrAspIAsrAspAlaIglYlPheIleYrSleuAsAlaLeuArGleu 398  
 1198 AAG 1200  
 399 Arg 399

```

seq_name: sp.Bacteria:Q9HY84
seq_documentation_block:
ID      Q9HY84      PRELIMINARY;      PRT;      405 AA.
AC      Q9HY84;
DT      01-MAR-2001 (TREMBLrel, 16, Created)
DT      01-MAR-2001 (TREMBLrel, 16, last sequence update)
DT      01-JUN-2001 (TREMBLrel, 17, last annotation update)
DE      ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
GN      LIGASE)
DN      ARGG OR PA3525.
OS      Pseudomonas aeruginosa.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC      Pseudomonas.
RX      NCBI_TaxID=287;
RP      SEQUENCE FROM N.A.
RC      STRAIN=PA01;
RX      MEDLINE=20437337; PubMed=10984043;
RA      Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA      Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA      Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA      Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardly K., Lim R.M.,
RA      Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA      Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT      "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT      opportunistic pathogen."
RL      Nature 406:959-964(2000).
CC      -I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC      PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC      -I- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC      PATHWAY.
CC      -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC      -I- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
DR      EMBL; AE004773; AAG69613.1; -.
DR      InterPro: IPR001518; Argininosuc_synth.
DR      Pfam: PF00764; Argininosuc_synth; 1.
DR      ProDom: PD003544; Argininosuc_synth; 1.
DR      PROSITE; PS00564; ARGININOSUCCIN SYN 1; 1.
DR      PROSITE; PS00565; ARGININOSUCCIN SYN 2; 1.
KW      ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW      Urea cycle.
SQ      SEQUENCE 405 AA; 45297 MW; 1C3DB39EB1869E5 CRC64;

alignment_scores:
Quality: 927.50      Length: 404
Ratio: 3.011      Gaps: 6
Percent Similarity: 76.238      Percent Identity: 47.030

alignment_block:
US-09-775-693-1 x Q9HY84      ..

```

Align seg 1/1 to: Q9HY84 from: 1 to: 405

```

1  ANGTCCAGCAAGAGCTCCGTGGTCTGGCCCTACAGTGGCGGCTGGACAC 50
1  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1  MetAlaAspValLysValValLeuAlaTyrSerGlyGlyLeuAsp 17
51  CTCGTGCATCCCTGCTGTGAGCA...CAAGCTTATGACGTGATG 97
51  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17  rSerValIleLeuLysTrpLeuLnsPthrTyrAsnCysGluValValT 34
98  CCTATGCTGCAACATTGGCCAGCAAGACTTCGAGCAAGCCAGAG 147
98  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34  hrPheThrAlaAspLeuGlyGlnGlyGluValAluProAlaArgAla 50
148  AAGGCACTGAGCTTGGGGCAAAAGGTTCATTGAGGATCTCAGCAG 197
148  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51  LysAlaArgAlaMetGlyValLysGluIleTyrIleAspAspLeuArg 67
198  GGAATTGTTGGAGAGTTCATCTGGCCGCCATCCAGCTCCAGCCACTG 247
198  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67  uGluPheValArgAspPheValTyrProMetPheArgAlaAsnThrValT 84
248  ATGAGAGCCGCTACCTCTGGCCACCTCTCTGCAAGGCCCTCAGTCCG 297
248  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
84  YrGluGlyGluTyrLeuLeuGlyThrSerIleAlaArgProLeuIleAla 100
298  CGCAACAGTGAATGCCCGCAGCGGAGGGGCCAGATGATGTGCCCA 347
298  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101  LysArgLeuIleGlyIleAlaAsnGluThrGlyAlaAspAlaIleSer 117
348  CGGGCCAGCAAGAGGGAAGCATGAGTCCGTTGAGCTGAGTGTCT 397
348  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
117  sGlyAlaThrGlyLysLysAsnAspGlnValArgPheGluLeuGlyAlaT 134
398  ACTGACTGGCCCCAGATAAAGTCAATGCTCCCTGAGAGATGCTGAA 447
398  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134  YrAlaLeuLysProGlyValLysValIleAlaProTprArgGluTprAsp 150
448  TTGTACAAACCGGTTCAAGGGCCGCAATGACTGATGAGTACGCAAGCA 497
448  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151  LeuLeuSerArgGluLys...LeuMetAspTyrAlaGluLys 163
498  ACAGGGATTCCTCATCCG...GTCACTCCCAAGACCCCTGGAGCA 541
498  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
163  ShisGlyIleProIleGluArgHisGlyLysLysSerProTyrSerM 180
542  TGGATGAGAACCTGACATGACATGACAGAGCTGAGATCTTGAGAAC 591
542  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180  eTAspAlaAsnLeuLeuHisIleSerTyrGluGlyValLeuGluAsp 196
592  CCCAAGAACCAAGCCCTCCAGGCTGTACAGCAAGACCCAGGACCCG 641
592  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
197  ThrTrpTrpGluHisGluGluAspMetTyrTrpTrpAlaSerProG 213
642  CAAGACCCCAACACCCCTGACATTCGTGAGATGAGTTCACAAAAGGG 691
642  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213  uAsnAlaProAspThrProThrTyrIleGluLeuThrTyrArgLysGly 230
692  TCCCTGTGAGGTGACCAACGTCAGATGGCAGCACCCAGCAAGCTCC 741
692  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230  sPrlaValAlaIle...AspGly...LysAspMetThrPro 241
742  TTGGAGCTTCATGATGACAGAGTCCGGGCAAGCAGAGGCTGGG 791
742  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242  AlaGluValLeuThrGluLeuAsnArgValGlyGlyIleAsnGlyIle 258
792  CCGTATGATCGTGGAGAACCGTTCATGGAATGAAGTCCGAGGTA 841
792  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
258  YrGluLeuAspIleValGluAsnArgTyrValGlyMetLysSerArgGly 275
842  TCTAGACAGCCCGCAGCAGCACCCTTACCATGCTCATTTAGACATC 891
842  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

275  YsTyrGluThrProGlyGlyThrIleMetLeuLysAlaHisArgAlaIle 291
892  GAGCGCTTCACCATGACCGGAAAGTGGCAAAATCAACAGAGCTGGG 941
892  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
292  GluSerIleThrLeuAspArgGluValAlaHisLeuLysAspGluLeu 308
942  CTGAAATTTGCTGAGCTGCTATACCGGTTTACGGCTTACCGCTGAGT 991
942  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
308  tProLysTyrAlaSerLeuIleTyrThrGlyTyrTrpPserProGlu 325
992  GTGAATTTGTCCGACATCGCCAGTCCAGGAGGAGGAGGAGGAGGAG 1041
992  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325  rGluMetLeuGlnIleMetIleAspAlaSerGlnValAlaAsnValS 341
1042  AAGTGCAGGTGCTCCCTCAGAGGCGCAGGTGATCATCTCGCCGCGGA 1091
1042  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
342  ValValaGluLeuLysLeuTyrLysGlnValValValaGlyArgLy 358
1092  GTCCCACTGTCTCTACATGAGAGCTGGTGCATGAAACGTGCAGG 1141
1092  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358  sSerAspAspSerLeuPheAspAlaAsnIleAlaThrPheGluGluAsp 375
1142  GTGAT...TATAGCCCACTGATGCCACCGGGTTCATCATCATTC 1188
1142  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375  YrGlyAlaTyrAsnGlnAlaAspAlaAlaGlyPheIleLysLeuAsn 391
1189  CTCAGCGTGAAG 1200
392  LeuArgMetArg 395
seq_name: sp_bacteria:Q9PHK7
seq_documentation_block:
ID Q9PHK7 PRELIMINARY; PRT; 406 AA.
AC Q9PHK7:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE ARGININOSUCCLINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
GN ARG OR C30665.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=NCCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE -> AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCLINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ARGININOSUCCLINATE SYNTHASE FAMILY.
CC EMBL: AL139075; CAB75297.1; -.
CC InterPro: IPR001518; Arginosuc_synth.
CC Pfam: PF00764; Arginosuc_synth; 1.
CC ProDom: PD003544; Arginosuc_synth; 1.
CC PROSITE: PS00564; ARGININOSUCCLIN SYN 1; 1.
CC PROSITE: PS00565; ARGININOSUCCLIN SYN 2; 1.
CC ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;
KW Urea cycle.
SO SEQUENCE 406 AA; 45578 MW; 8A1E137AF30EC77F CRC64;

```

alignment\_scores:                   Quality: 892.00                   Length: 397  
                                   Ratio: 3.014                   Gaps: 6  
                                   Percent Similarity: 74.559           Percent Identity: 47.355

alignment\_block:  
 US-09-775-693-1 x Q9PHK7                   ..

Align seg 1/1 to: Q9PHK7 from: 1 to: 406

```

19 GTGTTCTGGCCCTACAGTGGCGGCTGGACACTGCTGCTGCTG 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
8 ValValLeuAlaIyrSerIyGlyLeuAspThrSerIleLeuLys 24
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
69 GCTGAAG...GAACAAGGCTATGACGTCATGCTGCTGCTGCTG 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
24 pleuGlnaspGluTyrAsnGlyValValThrPheThrAlaAsp 41
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
116 GCCAAGAAGAGACTTGCAGAGAGCCAGAGAGAGAGAGAGAG 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
41 IyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGln 57
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
166 GCCAAAAG...GTGTTCAATGAGATGTCAGCAGGAGTTGTG 209
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
58 IleYsGlnGlnAsnIlePheIleIyAspLeuArgAspGluP 74
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
210 GGAGTTCATCTGGCCGCGCATCCAGTCAGCCAGCTGATGAG 259
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
74 SasprYrValPheProMetPheArgAlaAsnAlaIleYrGln 91
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
260 ACCTCTGGGACACTCTCTTGGCCAGCCCTGATGCGCCGCA 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
91 YrLeuLeuIyThrSerIleAlaIyArgProLeuIleAlaY 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
310 GAATGCGCCAGCGGGAGGGGCAAGTATGTCCTCCACGGCC 359
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
108 GlnIleAlaLeuGlnThrIyAlaAspAlaValSerHisGly 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
360 AAGGGGAGAGCATGAGTCCGTTGAGCTCAGTCTACACAGG 409
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
124 YrYsGlnAsnAspGlnValArgPheGlnLeuGlyIyLeuAl 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
410 CCCAATAAAGTCAATGCTGCTGCTGAGAGATGCTGAATTC 459
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
141 roAspLeuYsIleIleAlaProTrpArgGluTrpAspLeu 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
460 TTCAAGGGCCGCAATGACCTGATGAGTACCAAGCAACAG 509
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
158 GluYs.....LeuLeuAlaIyAlaGlnIyShIyGlyLea 170
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
510 CATCCCGTCACTCCC...AAGAACCCGTTGAGATGAGAGAC 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
170 PileSerIyYsIyGlyYsSerProIySerMetAspAlaSer 187
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
557 TGCACATGACGAGGCTGAGTCCGCGAGAACCCCAAGAAC 606
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
187 euHisIleSerIyGlnIyLeuValLeuGlnAspProAlaH 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
607 CCTCAGGCTCTACACGAGAGCCAGACCCCAAGAGCCCA 656
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
204 GluIuAspMetTrpArgTrpSerIySerProIyAspAla 220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
657 CCCGACATTTCTCGAGATCCAGTTCAAAAAAGGGCTCC 706
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
220 uSerGlnIleIleGlnLeuAspPheGlnIyGlyAspLeu 237
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
707 CCAAGTCAAGAGATGACACACCCAGCAACCTCTTGAGG 756
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
237 snGlyIyGlnYs.....LeuSerProAlaGlyLeuLeu 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
757 TACCTGACGAAGTGGCGGCAAGCATGGCTGGCGCTGAT 806
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

```

```

249 LysLeuAsnGlnLeuGlyCysIyShIyGlyIleGlyIyArgLeuAspIleVal 265
807 GGAGAACCGCTTCATTTGAGATGAGTCCGAGGATATACGAGACCCAG 856
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
265 IeIuAsnArgYrValIeGlyMetIySerArgIyCysYrGlnThrProG 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
857 CAGGACCATCTCTTACATGCTCATTTAGACATGAGCCCTCAGCATG 906
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
282 IyGlyThrIleLeuLeuYsAlaHisArgAlaLeuGlnSerIleThrLeu 298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
907 GACCGGAGAGTCCGCAAAATCAACAGAGCTGGGCTGAAATTTGCGA 956
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
299 AspArgGlnAlaAlaHisIleYsAspGluLeuMetProIyYrAlaSe 315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
957 GCTGCTATACCGGTTTACGCGCTAGCCCTGAGTGAATTTGCTCCGC 1006
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
315 rLeuIleYrAsnGlyYrTrpPheSerProGluArgMetLeuGlnA 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
1007 ACTGCATGCCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1056
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
332 lAlaLeuIeAspGlnSerGlnIleHisAlaAsnGlyArgValIyLeuGln 348
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
1057 GTCTCAAGGCGCAGGTGTACATCTCGCGCGGAGTCCACATG...TC 1103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
349 LeuYrIySgIyAsnValMetValIleGlyArgGlnSerAlaAsnAsp 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
1104 TCTCTACATAGAGAGCTGTGAGCATGAACGTGACAGGTGATTATGAGC 1153
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
365 rLeuPheAsnAlaAlaIyYrCysThrPheGlnGlnAspGluValYrAsn 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
1154 CAACGTATGCCACCGGTTATCATCAATCAATTCCTCCAG 1194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
382 lIyAspAlaAlaGlyPheIleYsIyLeuAsnAlaLeuArg 395
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

```

seq\_name: sp\_bacteria:09K820

seq\_documentation\_block:

ID	09K820	PRELIMINARY:	PRT:	409 AA.
AC	09K820:			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULINE--ASPARTATE LIGASE).			
GN	ARGG OR BH3187.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=86665;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C-125 / JCM 9153;			
RX	MEDLINE=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,			
RA	Horikoshi K.;			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RT	halodurans and genomic sequence comparison with Bacillus subtilis."			
CC	Nucleic Acids Res. 28:4317-4331(2000).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-CITRULINE + L-ASPARTATE = AMP +			
CC	PYROPHOSPHATE + L-ARGININOSUCCINATE.			
CC	-1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC			
CC	PATHWAY.			
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).			
CC	-1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.			
DR	EMBL: AP001518; BAB06906.1; -			
DR	InterPro: IPR001518; Arginosuc_synth.			
DR	Pfam: PF00764; Arginosuc_synth; 1.			
DR	ProDom: PD003544; Arginosuc_synth; 1.			
DR	PROSITE: PS00564; ARGININOSUCCIN_SYN_1; 1.			
DR	PROSITE: PS00565; ARGININOSUCCIN_SYN_2; 1.			
KW	ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;			
KW	urea cycle.			

```

757 TACGTACGAAAGTGGCGGGAAGCATGCGTGGCCGCTATTGACATCGT 806
242 GTCGAGAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
807 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
258 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
857 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
275 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
907 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
292 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
957 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
308 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
1007 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
325 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
342 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
1107 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
358 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
1157 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
375 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
1207 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
392 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856
seq_name: sp_bacteria:09K423

seq_documentation block:
ID 09K423 PRELIMINARY; PRT; 404 AA.
AC 09K423:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
LN LIGASE).
GN ARGG.
OS Moritella sp. 2674.
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Moritella.
NCBI_TaxID=111291;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=2674;
RX MEDLINE=20158877; PubMed=10692366;
RA Xu Y., Liang Z., Legrain C., Ruger H.J., Glansdorff N.;
RT "Evolution of Arginine biosynthesis in the bacterial domain: Novel
RT gene-enzyme relationships from psychrophilic moritella strains
RT (Vibrionaceae) and evolutionary significance of n-alpha-acetyl
RT ornithine."
RL J. Bacteriol. 182:1609-1615(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC -1- PROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGinine BIOSYNTHETIC
CC PATHWAY.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.
DR EMBL: AJ252020, CAB95017.1, -.
DR InterPro: IPR001518; Argininosuccin synthase.

```



DR Pfam: PF00764; Arginosuc\_synth. 1.  
 DR Prodom: PD003544; Arginosuc\_synth. 1.  
 DR PROSITE: PS00564; ARGININOSUCIN\_SYN\_1.  
 DR PROSITE: PS00565; ARGININOSUCIN\_SYN\_2.  
 DR AMP-binding: Arginine biosynthesis; Ligase; Urea cycle.  
 KW SEQUENCE 404 AA; 44291 MW; EBF8AFB4E87F20 CRC64;

## alignment\_scores:

Quality: 818.00 Length: 401  
 Ratio: 2.801 Gaps: 9  
 Percent Similarity: 72.818 Percent Identity: 44.638

## alignment\_block:

US-09-775-693-1 x Q9K423 ..

Align seg 1/1 to: Q9K423 from: 1 to: 404

```

19 GTGGTTCGAGCTACAGTGGCGCTGGACACCTGTCGATCCGCTGCTG 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 pLeuGlnGlu...AsnTyrAspAsnGlySerGluLeuValAlaPheValAla 42
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
69 GCTGAAGAGCAAGCTATGAC.....GTCATTCCTTCTGGGCA 109
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 pLeuGlnGlu...AsnTyrAspAsnGlySerGluLeuValAlaPheValAla 42
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 ACATTTGGCCAG...AAGGAAGACTTCGAGGAAGCCAGGAAGACGAC 156
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
42 spValGlyGlnGlyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 58
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 AAGCTTGGGCAAAAGGTTCATTTGAGGATGTGACAGGAGCTTGT 206
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 AlaserGlyAlaserGlyCysTyrValAlaPheLeuLysAspLeuVal 75
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 GGAGAGCTCATCTGCGCGGCAATCCAGTCACGACCTGATGAGAC 256
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
75 IgluAsnTyrLeuTyrProthreuleuSthrglyAlaValTyrGlnGly 92
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 GCTACTCTCTGGGCACTCTCTGCGAGCCCTGATCGCCCAACAA 306
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 hTyrLeuLeuGlyThrSerMetAlaArgProIleIleAlaLysAlaGln 108
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
307 GTGGAATTCGCCAGCGGAGGAGGCCAAGTATGTCGCCAGCGCCAC 356
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 ValGluIleAlaArgLysValGlyAlaAspAlaLeuGlySthGlyCys 125
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 AGGAAGGGAAGATCAGTCCGCTTGGAGCTGAC...TCAC 403
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
125 rGlyLysGlyAsnAspGlnIleArgPheGlu...SerCysPheAlaAla 141
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
404 TGCGCCCAAGATAAGTCAATGCTCCCTGGAGATGCCGATTTCTAC 453
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 euAlaProGluLeuThrValIleAlaProIleAlaLysPhePasp..... 155
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
454 AACCGGTCAAGGCGGCAATGACCTGTATGATGATCCCAAGACACG 503
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 .....LeuThrSerArgLysLeuLeuGlnGlnGlnGlnGlnGlnGln 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 GATTCCTCCATCCGCTACCTCCCAAGACCCGTCGAGATGAGAAC 553
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 pIleProIleAlaAlaSerAlaThrLysIleTyrSerIleArgAspAlaAsn 187
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 TCATGACATCAGTACGAGCTGGAATCTGGAGACCCCAAGAACCAA 603
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
187 latrPhisIleSerHisGlnGlyGlnGlnGlnGlnGlnGlnGlnGln 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
604 GCGCTCCAGCTCTACAGCAAGACCCAGACCCAGCAAGCCCAAG 653
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
204 ProSerLysGlnValTrpThrMetTrpValAspProIleAspAlaPro 220
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
654 CACCCCGACATTCGAGATGATCAAAAAGGGGCTCCCTGTGAAG 703
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 nGluProGluPheLeuThrIleSerValValLysGly...GluIleThr 236

```

```

704 TGACCAAGCTCAAGATGGACACCCAGCAAGCTCTTGAGACTCTTC 753
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
236 laValAsnGlyGlnAlaMetSerProIleArgPhe.....Leu 248
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
754 ATGATCTGACGAAGATCGCGGAGCAAGATGCGCTGGCCGATATGAC 803
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
249 MetTyrLeuAsnGlnLysAlaAlaIleIleGlyValGlyAlaLysP 265
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
804 CGTGAAGAACCGCTTCAATGATGAACTCCCGAGGATATCTAGACAG 853
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
265 euAlaGlnAsnArgLeuValGlyMetLysSerArgGlyCysTyrGln 282
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
854 CAGCAGGACCATCTTACCATGCTCATTTAGACATCGAGCCCTCAC 903
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
282 rGlyGlyThrValMetValGlnAlaLeuArgGlyIleGlnGlnLeu 298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
904 ATGACCGGGAAGTGGCAATCAACAGGCTGGGCTGAATTTGC 953
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
299 LeuAspLysThrThrArgLysTrpLysGlnThrAlaAlaGlnPhe 315
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
954 TGAGCTGCTGATACCGCTTACGGCTGAGCTGATGATTTGTC 1003
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
315 rHisLeuValTyrAspGlyArgTrpPheThrProLeuGlySalaSer 332
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1004 GCCACTGCATTCGCCAAGTCCAGAGCGAGTGAAGGAAGTGCAGTG 1053
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 euAlaAlaAlaGlyThrLeuAlaGlnGlnGlnGlnGlnGlnGlnGln 348
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1054 TCCGCTCTCAAGGCGGCAAGTCAATCCCGCGGAGATGCCACTGTC 1103
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
349 LysMetTyrGlySerValGlnAlaIleGlnLysLysSerProAsn 365
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1104 TCCTTACATGAGAGCTGTGTGACATGAACGTGAGGTGATTTGAC 1153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
365 rLeuTyrSerGlnGlnPheAlaThrPheGlyAspAsnValTyrAsp 382
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1154 CAAGTATGCCAGCGGCTTATCATCATCAATGATCCCTC..... 1197
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
382 InsertAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 398
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1198 AAG 1200
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
399 Lys 399

```

seq\_name: sp.plant:Q9SZX3

seq\_documentation\_block:

ID	Q9SZX3	PRELIMINARY:	PRT:	498 AA.
AC	Q9SZX3			
DT	01-MAY-2000 (TREMBlrel. 13, Created)			
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	ARGININOSUCCINATE SYNTHASE-LIKE PROTEIN (EC 6.3.4.5).			
GN	F617.40 OR AT4G24830.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Beyan M., Terryn N., Ardiles W., Buyschaert C., Dasseville R.,			
RA	De Clerck R., De Keyser A., Neyt P., Roupe P., Van Den Daele H.,			
RA	Villarroel R., Gielens J., Van Montagu M., Bancroft I., Hohelsel J.,			
RA	Mewes H.W., Mayer K.F.X., Schueller C.,			
RA	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;			
RA	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			



seq\_name: sp\_bacteria:Q9K4Y8

```
seq_documentation_block:
ID_Q9K4Y8;          PRELIMINARY;          PRT:          404 AA.
AC_Q9K4Y8;
DT_01-OCT-2000 (TREMBLrel. 15, Created)
DT_01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT_01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE_ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE_LIGASE).
CN_ARGG.
OC_Moritella sp. 2693.
OC_Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC_Moritella.
OX_NCBI_TaxID=111292;
RN_[1]
RP_SEQUENCE FROM N.A.
RC_STRAIN=2693;
RA_MEDLINE=20158877; PubMed=10692366;
RX_Xu Y., Liang Z., Legrain C., Kuger H.J., Glansdorff N.;
RT "Evolution of Arginine biosynthesis in the bacterial domain: Novel
RT gene-enzyme relationships from psychrophilic Moritella strains
RT (vibrionaceae) and evolutionary significance of n-alpha-acetyl
RT ornithinase.";
RL_J. Bacteriol. 182:1609-1615(2000).
CC_-I- CATALYTIC ACTIVITY: ATP + L-CITRULLINE + L-ASPARTATE = AMP +
CC_PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC_-I- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
```

CC PATHWAY.  
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 DR EMBL: AJ252021; CAB95023.1; -  
 DR InterPro: IPR001518; Arginosuc\_synth.  
 DR Pfam: PF00764; Arginosuc\_synth.1.  
 DR ProDom: PD003544; Arginosuc\_synth.1.  
 DR PROSITE: PS00564; ARGININOSUCCIN\_SYN\_1; 1.  
 DR PROSITE: PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
 KW ATP-binding; Arginine biosynthesis; Ligase; Urea cycle.  
 SQ SEQUENCE 404 AA; 44432 MW; 9ELC4D037AF5923F CRC64;

alignment\_scores:  
 Quality: 776.00 Length: 403  
 Ratio: 2.713 Gaps: 10  
 Percent Similarity: 70.968 Percent Identity: 43.672

alignment\_block:  
 US-09-775-693-1 x Q9K4Y8 ..

Align seg 1/1 to: Q9K4Y8 from: 1 to: 404

```

19 GTGGTTCTGGCCCTACAGTGGCGCCGTGACCTGCTGATCTCTGTC 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 ValValValValValValValValValValValValValValValVal 24
69 GCTGAAGACACAGCCTATGAC.....GTCATTGCCCTATCTGG 106
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
24 pleuGinglu...AsnTyrAspAsnCysGluIleValAlaLeuPheVal 40
107 CCAACATTGGCCAG...AAGAGACCTTCGAGAGACCGACGAGAGCA 153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 laaSPValGluGlnGluGluGluGluGluGluGluGluGluGluGlu 56
154 CTGAAGCTTGGCGCCAAAAGGTTCATTGAGCATGTCCAGGAGGATT 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 leuAlaSerGluAlaSerGluCysTyrValValAlaSerLeuLysAsp 73
204 TGTGAGAGTTCATCTGGCGCCGCTGACCTCCAGCGCCTGATGAGG 253
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 uValGluAsnTyrIleTyrProThrLeuLysThrGluValValVal 90
254 ACCGCTACCTCTGGGACCTCTCTGCGAGGCCCTGCATGCGCCGAA 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 lyThrTyrLeuLeuGluTyrSerMetAlaArgInsThrIleAlaLys 106
304 CAAGTGAATTCGCCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 353
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 GlnValGluIleAlaArgLysValGluValAlaSerPalaLeuLys 123
354 CACAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 400
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 sThrGluLysGluAsnSprInIleArgPheGlu...SerCysPheAla 139
401 CACGCGCCCGCCAGATTAAGTATTCCTCCGAGGAGGAGGAGGAGG 450
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 laLeuAlaProGluLeuThrValIleAlaProThrArgIleTyrAsp 154
451 TACAACGGTTCAAGGGCCGACATGACCTGATGAGTACGCCAAGACA 500
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 .....LeuThrSerAlaArgLysSerLeuLeuGluIleValAla 168
501 CGGGATTCCTCCATCCCGGTCACTCCCAAGAACCCGTGAGATGAGAG 550
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 gaSPileProThrAlaAlaSerGluTyrThrIleTyrSerAlaGasp 185
551 ACCTCATGCATCATCAGTACGAGGCTGGAATCTTGAGAACCCCAAGAC 600
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 smaIaTrpHisIleSerHisGluGluGluGluGluGluGluGluGlu 201
601 CAAGCGCTCCAGTCTTACACGAGAGACCGAGGCCAGGCCAAGCCCC 650
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

202 GlnProSerLysGluValIleThrMetThrValAlaSerProIleAspAla 218
651 CAACACCCCTGCATCTTCGATGAGTTCATCAAAAAGGGCTCCCTGCA 700
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 oAsnGluProGluPheLeuThrIleSerValValIleGly.....G 232
701 AGTCAACCAACGCTCAAGAGTGCACACCACACGACCTCTGTGAGCTC 750
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 luIleThrAlaValAlaSerGluGluGluMetLys.....LeuIleLea 246
751 TTCATGTACCTGACGAGTGGCGGCAAGCATGCGGCGCTGATGCA 800
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 TyrThrTyrLeuAsnGluLysAlaAlaAlaHisGluValGlyArgVala 263
801 CATGCTGAGAACCGCTTCATTGAAATGAATGCCGAGGCTTCACGAGA 850
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
263 pIleValGluAsnArgLeuValGluMetLysSerArgGlyCysTyrGlu 280
851 CCCCAGCAGCAGCATCTTACATGCTCATTTAGACATGAGCCCTTC 900
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 hrProGlyGlyThrValMetValGluValAlaLeuArgGlyIleGluGlu 296
901 ACATGACCGCGGAAGTGGCGCAAAATCAAAACAGCGCTTGAATT 950
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 ValLeuAspLysIleThrArgLysTyrPylHisThrValAlaAlaGlu 313
951 TGCTGACCTGCTATACCGGCTTACGCGCTACCGCTGAGTGAATTG 1000
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 eSerHisLeuValLysArgLysPylArgThrProLeuCysAlaSerL 330
1001 TCCGCCCATCGCATCGCAATCCAGAGCATGAGTGAAGGAAAGTGCAG 1050
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 euLeuAlaAlaAlaGlyThrLeuAlaGluGluMetAsnGlyValAla 346
1051 GTGTCCGCTCTCAAGGCGCAGGTGTACATCTCGCGCGGAGGCCACT 1100
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 ValLysMetLysGlySerValGlnAlaValGlnLysGlnSerProAs 363
1101 GTCTCTCTCAATGAGAGTGTGAGCATGAAGCTGCAGGCTGATTAT 1149
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 nSerLeuTyrSerGluGluPheAlaThrPheGlyAspAsnValTyrA 380
1150 ..GAGCCACTGATGCCACCGGTTTCATCAACATCAATTCCTTC 1191
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
380 spAspLysSerHisAlaGluGlyPheIleArgLeuTyrSerLeuSer 396
1192 AGGCTGAAG 1200
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 ArgIleLys 399

```

seq\_name: sp\_bacteria:Q89VC7

seq\_documentation\_block:

ID Q89VC7 PRELIMINARY; PRT; 401 AA.

AC Q89VC7;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE ARGININOSUCCINATE SYNTHASE.

GN ARGG OR SA0822.

OS Staphylococcus aureus subsp. aureus N315.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI\_TaxID=158879;

RN [1]

RP SEQUENCE FROM N.A.

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iino T., Kanemori M.,  
 RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Uji Y.,  
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,  
 RA Hirakawa H., Kohara S., Goto S., Yabuzaki J., Kanehisa M.,  
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
 RA Ogasawara N., Hayashi H., Hiramatsu K.;

660 TGACATTCCTCGAGAACGCAACGAAACGTTGGTGTG  
||| :: ||| |||||::|||||:::  
||| :: ||| |||||::|||||:::

```

710 AGCTAAGATGGACACACCACAGACCTCTTGAGCTCTTCAATGAC 759
      ||||| |||::: :: Aspleuleuthr 241
229 ..... AspglylvsThrTglnuLeuasp... Aspleuleuthr 241
      ||||| |||::: :: Aspleuleuthr 241
760 CTGAACGAAGTGGCGGGCGACAGCATGGCGGGCGCATTGACATCGTGA 809
      ||||| ::||| ||||| ||||| ||||| |||||
242 LeuAsnAlaLeuAlaGlyLysHisGlyIleGlyAlaHisIleAspHisValG1 258
      ||||| ::||| ||||| ||||| ||||| |||||
810 GAACGCGCTCATTTGGATGAATGATGCCGAGGTCTACAGACCCACAG 859
      ||||| ::||| ||||| ||||| ||||| |||||
258 uAsnArgLeuValGlyIleLysSerArgGluIleTglnAlaProAlaA 275
      ||||| ::||| ||||| ||||| ||||| |||||
910 CGGAGATGGCGCAAAATCAACAGACGCTGGAGCTTGAATTTGCGAGCT 959
      ||||| ::||| ||||| ||||| ||||| |||||
292 LysAspAlaAlaHisPheLysProIleIleGluLysGlnPheAlaGluG1 308
      ||||| ::||| ||||| ||||| ||||| |||||
960 GGTGTATACGGTTTACGGCTTACGCGCTTACAGTGTAAATTTGCGGCAC 1009
      ||||| ::||| ||||| ||||| ||||| |||||
308 nLeuTyrAsnGlyLeuThrPheSerProLeuThrAspSerLeuLysLeuP 325
      ||||| ::||| ||||| ||||| ||||| |||||
1010 GGATGGCCAGATCCACAGACGAGGTGAAGGAAAGTACGAGTGGCTGC 1059
      ||| ::||| ||||| ||||| ||||| |||||
325 heIleAspSerThrGlnTglnTyrValSerGlyAspAlaTglnLysLeu 341
      ||| ::||| ||||| ||||| ||||| |||||
1060 CTCAGGGCGCAAGTGTACATCCCTCGCGCGGAGTCCCACTGCTCTTA 1109
      ||||| ::||| ||||| ||||| ||||| |||||
342 PheLysGlyAsnAlaIleValAsnGlyAlaTyrLysSerProTyrThreLeu 358
      ||||| ::||| ||||| ||||| ||||| |||||
1110 CATTGAGGAGCTGTGAGCATGAACGTGCACGGTGTATTGAGGCCAAG 1159
      ||||| ::||| ||||| ||||| ||||| |||||
358 rAspGluLysLeuAlaIleThrTyrThrLysGluAspAlaPheAsnGlnAspA 375
      ||||| ::||| ||||| ||||| ||||| |||||
1160 ATGCCACCGGTTTCATCAACATCAATTCGCTC 1191
      ||||| ::||| ||||| ||||| ||||| |||||
375 laaIaValGlyPheIleAspIleTyrGlyLeu 385

seq_name: sp_bacteria:Q9KNR8

seq_documentation_block:
ID Q9KNR8 PRELIMINARY: PRT: 404 AA.
AC Q9KNR8;
DI 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DI 01-OCT-2000 (TrEMBLrel. 17, Last annotation update)
DI 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE
DE LIGASE).
GN VC2642.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC Bacteria; taxid:666;
OX NCBI_taxid:666;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10932301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emdinova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonaId L., Utterback T., Fleischmann R.D., Nleman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nucleu 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: Asp + L-CITRULLINE + L-ASPARTATE = AMP +
CC PYROPHOSPHATE + L-ARGININOSUCCINATE.
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC
CC PATHWAY.

```

CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE ARGININOSUCCINATE SYNTHASE FAMILY.  
 DR EMBL: AE004330; AAF5783.1;  
 DR TIGR: VC2642;  
 DR InterPro: IPR001518; Argininosuc\_synth.  
 DR Pfam: PF00764; Argininosuc\_synth; 1.  
 DR ProDom: PD003544; Argininosuc\_synth; 1.  
 DR PROSITE: PS00564; ARGININOSUCCIN\_SYN\_1; 1.  
 DR PROSITE: PS00565; ARGININOSUCCIN\_SYN\_2; 1.  
 KW ATP-binding; Arginine biosynthesis; Complete proteome; Ligase;  
 KW urea cycle.  
 SEQUENCE 404 AA; 44465 MW; A65C73B442B5AF82 CRC64;

alignment\_scores:  
 Quality: 751.50 Length: 394  
 Ratio: 2.674 Gaps: 6  
 Percent Similarity: 71.320 Percent Identity: 41.878

alignment\_block:  
 US-09-775-693-1 x 09KNT8 ..

Align seg 1/1 to: 09KNT8 from: 1 to: 404

```

19 GTGGTCTGCTACAGTGGCGGCTGAGACCTGTGATCTGCTGTG 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
9  ValValValAlaValTyrSerGlyLeuAspThrSerValIleIlePro 25
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 GCTGAAGAGAACAGGCTATGAC.....GTATTGCTATCTGGCCACA 112
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 PleuLysGlu.....AsnTyrAspCysGluValAlaIleValAlaAsp 41
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 TTGGCCAG...AAGAAGACTTTCAGAGAACCCAGAGAACGAGCAAG 159
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 AlGlyGlnGlyAspGlnGluLeuLysGlyValGluAlaLysAlaLeuSer 57
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 CTGGGGCCAAAGAGTGTTCATGAGATGATGACGAGGAGCTTTGGA 209
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 SerGlyAlaSerGlnCysTyrIleValAlaPheLysGlnLupheVally 74
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 GGAGTTCATCTGGCGGCGCATCCAGTCCAGGCACTGATGAGACCGGT 259
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 SGLTyrIleTyrProThrLeuLysThrGlyAlaTyrTyrGlnGlyLys 91
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 ACCCTCGGCGACCTCTTGGCCAGGCGCTGCATGCGCCCGGCAACA 309
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 YrLeuLeuGlyThrSerMetAlaArgProValIleAlaLysAlaGlnVal 107
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 GAATGCGCCAGCGGAGGGGCCAAGTATGTGTCCAGCGGCGCACAGG 359
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 GluIleAlaArgLysValGlyAlaLysAlaLeuAlaHisGlyCysThr 124
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 AAGGGGAGACGATCAGTCCGCTTGGAGCTCAGCTGCTACCTGCTG 409
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 YrLysGlyAsnAspGlnValArgPheGlnGlyAlaPheAlaLeuVal 141
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 CCCAGATAAAGTCTATGCTCCCGGAGAGATGCTGATTTCTCAACCG 459
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 rAspLysHisValIleAlaIleAlaProTyrPargLupTyrPasp..... 153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 TTCAAGGGCCCAATGACCTGATGAGATAGCAAGACACAGCGGTTTC 509
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 LeuArGSerArGlnAlaCysLeuAspTyrLeuAlaIleAlaGlnAr 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 CATCCCGGTCATCCCAAGACCCGCTGAGACATGATGAGAACCTCAT 559
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 ocysAlaIleSerLeuThrLysIleTyrSerArGAspAlaAsnAla 187
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 ACATACGACTACGAGCTGCAATCTGAGAACCCCAAGACCAAGCGCT 609
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 lIsvAlSerThrGlnGlyValLeuGlnSerThrTTPAsnAlaProAsn 203
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 CCAGTCTCTACACGAAGCCAGACCCAGCAAGGCCCAAGCCCAACGCC 659

```

```

204 GluAspCysTrpValIleThrValAlaAspProGlnGlnAlaIleProAsnGlnAl 220
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 TGACATTCCTCGACATCGATGATTCAAAAAGAGCGCTGGAAGTGACCA 709
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 agLTrValIleThrLeuGlnValAlaHisIleGlyValIleAspIleValG 265
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 ACGTCAAGATGAGCACCACCACAGACCTCTTGGAGCTCTCATGATAC 759
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 .....AspGlyGluAla...MetThrProTyrAsnAlaLeuLeuTyr 248
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 CTGACGAAGCTGCGCGGCAAGCATGCGTGGCGGCTGATGACATCTGGA 809
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 LeuAsnGlnLysGlyAlaLysHisGlyValIleThrIleAspIleValG 265
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
810 GAACGCTTCATTTGGAATGATCCCGAGGTATCTACAGAACCCAGCAG 859
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 uAsnArGLeuValGlyMetLysSerArGlyCysTyrGluThrProGlyC 282
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 GCACCATCTTACCATGCTCATTTAGACATCGAGCGCTTCACACATGAC 909
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 lYrIleIleMetGlnAlaLeuAlaArgAlaValGlnIleuValLeuAsp 298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
910 CGGAAAGTGCACAAATCAACAAGCGCTGGAATTTGCTGACCT 959
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 LysThrSerPheGlnPheArgGlnGluLeuGlyIleLysAlaSerHisLe 315
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 GGTGATACGGGTTTACGGCTAGCCCTGAGTGAATTTGCGCCACT 1009
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 uValTyrAspGlyArgTyrPheThrProLeuArgGlnAlaValPheAla 332
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1010 GCATCCGCCAAGTCCAGAGCGAGCGAGGAAGAAAGTGCAGTCCGTC 1059
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 lAlaAspGlnLeuAlaLysAspValAsnGlyGluValAlaIleLysLeu 348
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1060 CTGAAGCGCCAGGTGATCATCTCGCGGAGAGTCCCGCTGCTGCTGA 1109
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 TyrLysGlyGlnAlaValAlaThrGlnLysArgSerAlaAsnSerLeuTyr 365
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 CATATGAGAGCTGTGATGACATGAACTGACGGGATATGAGCCCAAGTC 1159
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 rSerLysAspPheAlaThrPheGlyAlaAspGluValTyrAspHisSer 382
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1160 ATGCCACGCGGTTCATCATCATCATCATTCCTC 1191
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
382 lAlaGlyGlyPheIleAlaGlyLeuTyrSerLeu 392
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: sp\_bacteria:09CC10

seq\_documentation\_block:

ID	09CC10	PRELIMINARY	FR	399 AA
AC	09CC10			
DT	01-JUN-2001 (TREMblrel. 17, Created)			
DT	01-JUN-2001 (TREMblrel. 17, Last sequence update)			
DE	01-JUN-2001 (TREMblrel. 17, Last annotation update)			
OS	ARGININOSUCCINATE SYNTHASE.			
OC	MycoBacterium leprae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;			
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1769;			
GN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-TN;			
RA	MEDLINE=21128732; PubMed=11234002;			
RA	Cole S.T., Eglmeier K., Garnier T., James K.D., Thomson N.R.,			
RA	Wheeler P.R., Honore N., Churruarin C., Harris D.,			
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,			
RA	Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,			
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,			
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,			
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,			

RA Barrell B.G.; "Reactive gene decay in the leprosy bacillus."

alignment_scores:	685.50	Length:	414
Quality:	2.422	Gaps:	8
Ratio:	2.422	Percent Identity:	37.440
Percent Similarity:	68.357		

Align seg 1/1 to: 09CC10 from: 1 to: 399

[illegible]

```

648 CCCGACACCCCTGACATCTCTGACATGCGAGTGCACAAAAGGCGCTTGG 637
      |||||
      : : : : : |||||
212 ntrpAstrhrproAserbgluValleValgluYrheglnslgValbProv 229
      : : : : : |||||
698 TGAAGATGACCAAGTCAGATGGCCACACCCACCGACGACTCTTGGAG 747
      : : : : : |||||
229 alserlle.....AspGlySer..ProValSerMetIleuGlyu 241
      : : : : : |||||
748 CTTCTCATGTACTGACGAAAGTGCAGGCGAGCAAGAGTGGCGCTGGCCGAT 797
      : : : : : |||||
241 lalleluuValaLeuValnArgdValaValaValaValaValaValaVala 257
      : : : : : |||||
798 TGACATCTGTGAGAGAACGCTTCATGTGGAAATGAAGTCCCGAGATATCAG 847
      : : : : : |||||
257 uAspValValGlnuSpruValuValuIleuValuSerTrpValuIleuVal 274
      : : : : : |||||
848 AGACCCGACGAGGACCATCTTTACCATGCTCATTTAGACATTCAGAGGCC 897
      : : : : : |||||
274 luAlaProgluAlaMetuValuIleuIleuThrAlaHisIaValuIleuInd 290
      : : : : : |||||
898 TTGCACATGACCGGAGATGTCGCGAAATCAACAAAGGCGCTGGCGCTTGA 947
      : : : : : |||||
291 ValThrIleuGluValuArgIleuValuArgPheValuArgIleuValuThr 307
      : : : : : |||||
948 ATTTCCTAGCTGTGTATACGCGTTTCAGGCTACCCCTGAGTGTGAT 997
      : : : : : |||||
307 gtrPalaGluValuValuIleuValuIleuTrpValuSerProValuValuTha 324
      : : : : : |||||
998 TTGTCCGCACTGATCGCCCAATGCCAGGACGAGTGGAGGAAAGTG 1047
      : : : : : |||||
324 lalenuIuSerPheValaAlaIleuAlaThrGlnIleuValuIleuValuVal 340
      : : : : : |||||
1048 CAGGTGCTGCTTCAGGCGCAGCTGTACATTCGTCGCGCGGAGATCCCC 1097
      : : : : : |||||
341 ArgMetValuIleuValuIleuValuIleuValuValaValaValuValuVal 357
      : : : : : |||||
1098 ACTGTCTCTCAATGATGAGAGCTGTGTACATGAAACGTCAGCGGTGAT 1147
      : : : : : |||||
357 aGluIleuValuIleuValuIleuValuIleuValuIleuValuIleuValu 374
      : : : : : |||||
1148 ATGAGCCAACTGATGCCACCGGGTTCATCAACATCAATTCCTCAGCGTG 1197
      : : : : : |||||
374 heAspGlnSerAlaAlaAlaArgGlyPheValuValuValuValuValuVal 1236
      : : : : : |||||
1198 AAGGATATCATCTGCTTCAGGACCAAGTCACCTGGCCAA 1286
      : : : : : |||||
386 .....TyrGlyIleuProSerIleuAlaAlaAlaArg 395
      : : : : : |||||

seq_name: sp.archaea:09UX31

seq_domenclature_block:
ID 09UX31 PRELIMINARY; PRT; 406 AA.
DC 09UX31;
DT 01-MAY-2000 (TREMBLrel_13, created)
DT 01-MAY-2000 (TREMBLrel_13, last sequence update)
DT 01-JUN-2001 (TREMBLrel_11, last annotation update)
DE ARGININOSUCINATE SYNTHASE (EC 6.3.4.5).
GN Sulfolobus solfataricus; Sulfolobaceae; Sulfolobus.
OS Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OC NCBI_TaxID-2287;
OX 11
RN SEQUENCE FROM N.A.
RC STRAIN-DSM 1617 / P2;
RA Charlebois R.L., Strub R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,
RA Canteloni L.F., Curtis B., Duguet M., Eruso G., Fanny D.,
RA Gasterlandier T., Garrett R.A., Gordon P., Jeffries A.C., She O.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.D., Doolittle W.F.,
RA St Jean A., Sensen C.W.;
RA Ragan M.A.;
RT "gene content and organization of a 281-kbp contig from the genome of
the extremely thermophilic archaeon, Sulfolobus solfataricus P2."

```

RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y18930; CAB57663.1; -  
 DR InterPro: IPR001518; Argininosuc synth.  
 DR Pfam: PF00764; Argininosuc synth\_1.  
 DR ProDom: PD003544; Argininosuc synth\_1.  
 DR PROSITE: PS00564; ARGININOSUCIN\_SYN\_1; UNKNOWN\_1.  
 KW PROSITE: PS00564; ARGININOSUCIN\_SYN\_2; 1.  
 LIGASE.  
 SEQUENCE 406 AA: 46285 MW: 3956BAB9CF04C12 CRC64;

## alignment\_scores:

Quality: 656.50 Length: 394  
 Ratio: 2.441 Gaps: 9  
 Percent Similarity: 68.274 Percent Identity: 39.594

## alignment\_block:

US-09-775-693-1 x Q9UX31

Align seq 1/1 to: Q9UX31 from: 1 to: 406

```

19 GTGTTCTGCGCTACAGTGGCGGCTGGACCTCGCATCGCTGCTG 68
   ::::::::::::::::::::::::::::::::::::::::::::
18 ILevalLeuAlaTyrSerGlyGlyLeuAspThrValSerThrLeu 68
   ::::::::::::::::::::::::::::::::::::::::::::
69 GCTGAAGCAACA...GGCTATGACGTCATTCGCTATGGCCACATG 34
   ::::::::::::::::::::::::::::::::::::::::::::
34 pleuysglutrhphelysalaagluilethrvalthrvalaspar 51
   ::::::::::::::::::::::::::::::::::::::::::::
116 GCCAAGAGAAAGACTTGGAGAGCCAGAGAGAGAGAGAGAGAGAG 165
   ::::::::::::::::::::::::::::::::::::::::::::
51 LyglnLysaspaspRheLysLysIleGluLysArgLysLysLys 165
   ::::::::::::::::::::::::::::::::::::::::::::
166 GCCAAAGAGTTCATTCAGTGGATGTCAGAGAGAGAGAGAGAGT 67
   ::::::::::::::::::::::::::::::::::::::::::::
68 AlaserLysHisThrThrIleAspAlaValArgLysPheAlaSer 215
   ::::::::::::::::::::::::::::::::::::::::::::
216 CATCTGCGCGCCATCCAGTCCAGTCTATGAGAGAGAGAGAGAG 84
   ::::::::::::::::::::::::::::::::::::::::::::
84 rleAlaTyrAlaIleLysLeuAsnGlyLeuTyrGluGlyValTyr 101
   ::::::::::::::::::::::::::::::::::::::::::::
266 TGGGACCTCTCTTCCAGGCGCTGTCATGCGCCGCAAGAGAGT 101
   ::::::::::::::::::::::::::::::::::::::::::::
101 elserThrAlaLeuAlaArgProLeuIleAlaLysValAlaGlu 117
   ::::::::::::::::::::::::::::::::::::::::::::
316 GCCAGAGGAGAGGAGGAGTATGTGTCCAGGCGCCAGAGAGAG 365
   ::::::::::::::::::::::::::::::::::::::::::::
118 AlalysLysGluGlyAlaGluAlaValAlaIleLysSerThr 134
   ::::::::::::::::::::::::::::::::::::::::::::
366 GAAGATCAGTCCGCTTGGCTGAGTCTGCTACGCTGCTGCTGCT 415
   ::::::::::::::::::::::::::::::::::::::::::::
134 YasnAspGluValArgPheAspLeuAlaValLysAlaLeuTyr 151
   ::::::::::::::::::::::::::::::::::::::::::::
416 TAAAGTCAATGCTGCC...TGGAGATGCTGTAATCTCAAC 456
   ::::::::::::::::::::::::::::::::::::::::::::
151 alLysIleIleAlaProAlaArgLysIleTrpAsnMetThr 163
   ::::::::::::::::::::::::::::::::::::::::::::
457 CGGTTCAAGGCGCGCAATGACCTGATGAGTACCAAGAGAGAG 506
   ::::::::::::::::::::::::::::::::::::::::::::
164 .....ArgGluAspGluIleLysTyrAlaLysGluGlyVal 176
   ::::::::::::::::::::::::::::::::::::::::::::
507 TCCCATCCCGGCTACTCCCAAGAACCGTGGACATGAGTGAAC 556
   ::::::::::::::::::::::::::::::::::::::::::::
176 eptolLysVal...GluSerAspLysTyrSerThrLeuAspLys 192
   ::::::::::::::::::::::::::::::::::::::::::::
557 TGCATCAGTCACTGAGGCGGATTCCTGGAGAACCCCAAGAGAG 606
   ::::::::::::::::::::::::::::::::::::::::::::
192 rpeLysSerThrLeuGlyAspLysIleSerAspProSerLeuVal 208
   ::::::::::::::::::::::::::::::::::::::::::::
607 CTCACAGGCTCTACAGAGAACCCAGAGAGAGAGAGAGAGAGAG 208
   ::::::::::::::::::::::::::::::::::::::::::::
209 ProGluAspAlaPheGluTyrThrLys...GlnIleTyrAsn 222
   ::::::::::::::::::::::::::::::::::::::::::::

```

```

657 CCTGACATTCGAGATCGAGTTCACAAAAGGGGCTCCGTGAGAGTGA 706
   ::::::::::::::::::::::::::::::::::::::::::::
222 sLysGluIleValSerIleGluPheSerAsnGlyValProThrAlaVal 239
   ::::::::::::::::::::::::::::::::::::::::::::
707 CCACGTCAGAGATGGACCAACCAGACCTCTCTGGAGCTCTTCATG 756
   ::::::::::::::::::::::::::::::::::::::::::::
239 sngLysLys.....MetGluLeuAsnLysLeuValAsp 250
   ::::::::::::::::::::::::::::::::::::::::::::
757 TACCTGAGACAGAGTCCGCGCAGCATGGCGGCGCTATTCATCGT 806
   ::::::::::::::::::::::::::::::::::::::::::::
251 PheLeuAsnLeuLysPheGlySerHisGlyPheGlyArgValGluHis 267
   ::::::::::::::::::::::::::::::::::::::::::::
807 GGAGAACCGCTTCATTCGATGATGATCCGAGTATCAGAGAGAGAG 856
   ::::::::::::::::::::::::::::::::::::::::::::
267 egluAsnArgValAlaGlyPheLysSerArgGluValTyrGluValPro 284
   ::::::::::::::::::::::::::::::::::::::::::::
857 CAGGCACATCTCTTACATGCTCATTCATGACATCGAGGCTTCAC 906
   ::::::::::::::::::::::::::::::::::::::::::::
284 lAlaIleuGlyLeuIleTyrAlaHisIleAspLeuGlyThrIleTyr 300
   ::::::::::::::::::::::::::::::::::::::::::::
907 GACCGGAGAGTGGCAAAATCAACAGGCTTGGCTGATTCATGCTGA 956
   ::::::::::::::::::::::::::::::::::::::::::::
301 ThrProMetGluLeuAlaGlyPheLysArgHisIleAspGluLeuTyr 956
   ::::::::::::::::::::::::::::::::::::::::::::
957 GCTGCTGATACCGCTTTCACGCGCTACCGCTGAGTGAATTCCTG 1006
   ::::::::::::::::::::::::::::::::::::::::::::
317 pleuValTyrGluGlyLeuTyrPheGluProLeuArgGluThrLeuHis 1056
   ::::::::::::::::::::::::::::::::::::::::::::
1007 ACCTGATCGCCAGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGT 1056
   ::::::::::::::::::::::::::::::::::::::::::::
334 yValAlaAspGluMetAsnLysTyrPheGlyGluAlaLysValGlu 1106
   ::::::::::::::::::::::::::::::::::::::::::::
1057 GTCTCTAAGAGGCGGAGTACATCTCGCGGAGAGAGAGAGAGAGT 1106
   ::::::::::::::::::::::::::::::::::::::::::::
351 ValSerAsnGlySerPheArgLysIleValAlaGlyLysSerGlyTyr 367
   ::::::::::::::::::::::::::::::::::::::::::::
1107 CTACATGATGAGAGTGTGTCAGATGAACGTCGAGTGTATGAGAG 1156
   ::::::::::::::::::::::::::::::::::::::::::::
367 oLysSerLysIleAlaSerTyrAsn...LysGlyTyrTyr...Pro 382
   ::::::::::::::::::::::::::::::::::::::::::::
1157 CTGAT.....GCCACGCGGTTCATCAACATC 1182
   ::::::::::::::::::::::::::::::::::::::::::::
382 eAspGluMetAlaArgGlyPheIleGluIle 392
   ::::::::::::::::::::::::::::::::::::::::::::

```

seq\_name: sp\_human:043348

## seq\_documentation\_block:

ID 043348 PRELIMINARY; PRT: 166 AA.  
 AC 043348;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DE HYPOHETICAL PROTEIN R6007015.1 IN CHROMOSOME 7Q31 (FRAGMENT).  
 GN R6007015.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1] NCBI\_Taxid-9606;  
 RP SEQUENCE FROM N.A.  
 RL Stoenking T., Langston Y., Ahrens C.;  
 DR EMBL: AC003989; AAB96328.1; -  
 DR InterPro: IPR001518; Argininosuc synth.  
 DR Pfam: PF00764; Argininosuc synth.  
 DR ProDom: PD003544; Argininosuc synth; 2.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 166 AA: 18967 MW: 2666A3ADCFFBFFAD CRC64;

## alignment\_scores:

Quality: 604.50 Length: 185

Ratio: 4.003 Gaps: 2  
Percent Similarity: 81.622 Percent Identity: 67.568  
alignment\_block:  
US-09-775-693-1 x 043348 ..  
Align seg 1/1 to: 043348 from: 1 to: 166

```

682 MAAAAAGGGGTCCTGTGAGAGTGAACGTCAGAGATGACACCCCA 731
1 LysLysGlyValProValLysValThr...lleYsMetAlaProPonH 16
732 CCAGACCTCTGTGAGAGTGAACGTCAGAGATGACACCCCA 781
16 sctlnhrSerLeuGlnLeuPheValTyrLeuAsnGlnValThrGlyLysH 33
782 ATGGCGTGGCGCGATTTGACATCGTGAACCGCTTCATGGAATGAG 831
33 lsglYmeGlyCysIleAspIleMetGluAsnCysPheIleArgIleLys 49
832 TCCGAGGTATCTACGAGACCCGACAGACGACATCTTACATGCTCA 881
50 SerArgGlyIleTyrLysThrProAlaGlyThrIleLeuGlnAlaAs 66
882 TTTAGACATCGAGCGCTTCACCATGAGACCGGGAAGTGGCAAAATCAAC 931
66 PleuAspIleGlnAspPheThrMetAspArgGlnValHisLysIle... 81
932 AAGCGCTGGCTTGAATTTGCTGAGCTGGTGTATACCGGTTTACGGCT 981
81 ..... 81
982 AGCCCTGAGTGAATTTGCGCGCATGATGATGATGATGATGATGATGAT 1031
82 AsnProGlnCysLysSerValHisHisCysIleThrLysSerGlnLug 98
1032 AGTGAAGGGAAGTGCAGGTGTCGCTGCTGAGAGCGGCGAGTGTATATCC 1081
98 nmeGlnGlyLysValGlnValProIlePheLysGlyGlnValHisLys 115
1082 TCGCGCGGAGTCCCGCATGCTGTCTTACATGAGAGCGTGTAGCATG 1131
115 eusSerGlnLysSerLeuLeuSerLeuTyrLysGlnLugLeuValSerVal 131
1132 AACGTGAGGAGTATGATGAGACCATGATGATGATGATGATGATGATGAT 1181
132 AsnValGlnLysnTyrGlnProIleAspAlaThrGlyPheIleAsn 148
1182 CAATTCCTGAGCGTGAAGGATATGATGATGATGATGATGATGATGATGAT 1231
148 easnTyrLeuArgLeuLysGlnTyrHisTyrPheGlnSerLysValThr 165
1232 CCAAA 1236
165 alyys 166
seq_name: sp_archaea:09HMQ2
seq_documentation_block:
AC 09HMQ2 PRELIMINARY; PRT: 396 AA.
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE ARGININOSUCCLINATE SYNTHETASE.
GN ARGG OR VNG2437G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea: Euryarchaeota: Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCB1:TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=20504483; PubMed=11016950;

```

```

RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angelone C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Love T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005124; AGS20519.1;
DR InterPro: IPR001518; Arginosuc-synth.
DR Pfam: PF00764; Arginosuc-synth.
DR ProDom: PD003544; Arginosuc-synth.
DR PROSITE: PS00564; ARGININOSUCCLIN_SYN_1; UNKNOWN_1.
DR PROSITE: PS00565; ARGININOSUCCLIN_SYN_2; 1.
KW Complete proteome.
SQ SEQUENCE 396 AA; 41994 MW; 8D48C2DCAD94D092 CRC64;

```

alignment\_scores:

Quality: 539.50 Length: 424  
Ratio: 2.107 Gaps: 14  
Percent Similarity: 60.377 Percent Identity: 33.255

alignment\_block:  
US-09-775-693-1 x 09HMQ2 ..

Align seg 1/1 to: 09HMQ2 from: 1 to: 396

```

13 GGCTCCGTGTTCTGCTACAGTGGCGGCTGAGACCTG...TGCAT 59
4 GlyThrValAlaLeuAlaPheSerGlyLysPheThrValLysVal 20
60 CCTGCTGCTGCTGAGACAGGCTAT...GACGTCATTCCTATCTGG 106
20 LProLeuLysGlnLugTyrGlyTyrAspAspValIleGlyLthr 37
107 CCACATTTGGCGGCAAGGAA...GACTGAGAGACCCGAGAAAGGCA 153
37 AlaAspValGlyGlnProAspAlaGlnPheAlaAlaAlaArgLthrAla 53
154 CTGAGCTTGGGCGCAAAAGTGTTCATTTGAGATGATGAGAGGAT 203
54 AlaAlaLeuGlyValGlnHis...HisValAlaAspAlaThrAlaGln 69
204 TGTGAGAGAGTTCATCTGCGCGGCGATCCAGTCCAGGCGATGAGG 253
69 eAla...AlaLeuCysPheAspAlaValArgAlaAsnAlaThrTyrGln 85
254 ACCGTACTCTCTGGGCGACCTCTGTCGAGGCGCTGATGCGCGCAAA 303
85 Lys...TyrProLeuGlyThrAlaLeuAlaArgProValIleAlaAspAla 100
304 CAAAGTGAATTCGCCCGGAGGAGGCGCAAGTATGTGTCCACGCGGC 353
101 IleLeuSerValAlaGlnLugLysAlaAlaAlaLeuAlaHisGlyCys 117
354 CACAGGAAAGGAGGAGATGATGATGATGATGATGATGATGATGATGAT 403
117 sThrGlyLysGlyAsnAspGlnLeuArgPheGln... 128
404 TGGCCCCCAGATAAAGTTCCTGCTGAGGAGATGCTGAAATTCAG 453
129 .....AlaValTTPArgAlaSerAsp..... 135
454 AACCGTTTCAAGGCGCGCATGACTG..... 480
136 .....HisAspValCysAlaProValArgGlnLeu 146
481 .....ATGAGTACGCAAGACAGACAGGATTCGCA 511
146 yLeuThrArgLugLthrGlnLugLthrAlaAlaGlnLugLthrAspLeuPro 163

```



[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:44:27 ; Search time 1806.06 Seconds  
(without alignments)  
11317.456 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239  
Sequence: 1 atgtccagcaagagcttcgt.....gcaaggtcactgcgaataag 1239

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hgo\_hum:\*  
31: em\_hgo\_inv:\*  
32: em\_hgo\_rod:\*  
33: em\_hgt\_hum:\*  
34: em\_hgt\_inv:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	100.0	1560	9	HSASD
2	1232.6	99.5	1975	9	AK027126
3	1231	99.4	1595	9	BC009243
4	1093	88.2	180838	2	AC027153
5	1091.4	88.1	165439	2	AC025175
6	1067.8	86.2	1398	4	BOVASS
7	1058	85.4	188356	2	AC009550
8	1058	85.4	190665	2	AP003057
9	1058	85.4	201989	2	AC073169
10	1057.8	85.4	155019	2	AL353717
11	1055.4	85.2	159687	2	AC087311
12	1055.4	85.2	204371	2	AC026357
13	1052.8	85.0	107314	9	AC005000
14	1052.2	84.9	1868	9	HUMAS3PS
15	1032	83.3	110716	2	AC008717
16	1031.8	83.3	93898	2	AC078791
17	1030.4	83.2	197429	2	AC023666
18	1030	83.1	58408	9	AL160400
19	1030	83.1	151833	2	AC016875
20	1027.8	83.0	1495	10	RATASSA
21	1027.8	83.0	1495	10	RATASSA
22	1016.6	82.1	1561	10	MUSAS5B
23	1016.6	82.1	1645	10	MUSAS5B
24	1014	81.8	1877	9	HUMASALPS
25	1011.2	81.6	149752	2	AC004616
26	1009.6	81.5	160658	2	AC073582
27	988.2	79.8	113799	2	AC012502
28	988.2	79.8	173690	2	AC011307
29	986.2	79.6	176948	2	AL354700
30	986.2	79.6	196376	2	AL353705
31	961	77.6	194389	2	AC018693
32	959.4	77.4	347974	2	AC092833
33	570.8	46.1	132930	2	AC079046
34	567	45.8	81346	2	AC021496
35	551	44.5	808	9	HUMAS08
36	536.4	43.3	179658	2	AC024525
37	427.4	34.5	10007	1	AE001816
38	419.4	33.8	81346	2	AC021496
39	362.8	29.3	118103	9	AC003989
40	325.6	26.3	10040	1	AE001924
41	321	25.9	336	1	HUMAS05
42	315	25.4	10138	1	AE005687
43	310.4	25.1	360	6	AX192950
44	305.2	24.6	405	9	HUMD1A06M5
45	301.6	24.3	41255	8	SPBC428

## ALIGNMENTS

Result	1
HSASD	HSASD
LOCUS	LOCUS
DEFINITION	Human mRNA for argininosuccinate synthetase.
ACCESSION	X01630
VERSION	X01630.1 GI:28871
KEYWORDS	synthetase.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 1560)
AUTHORS	Bock, H.G., Su, T.S., O'Brien, W.E. and Beaudet, A.L.
TITLE	Sequence for human argininosuccinate synthetase cDNA
JOURNAL	Nucleic Acids Res. 11 (18), 6505-6512 (1983)
MEDLINE	84015386
FEATURES	location/Qualifiers
SOURCE	1..1560
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

```

misc_feature      40..48
                  /note="three tandem arginin codons"
CDS               76..1314
                  /note="argininosuccinate synthetase (aa 1-412)"
                  /codon_start=1
                  /protein_id="CAA25771.1"
                  /db_xref="GI:28872"
                  /db_xref="SWISS-PROT:P00966"
                  /translation="MSKSGSVVLAISGGLDSCILWLMEQGYDIYAVIANTIQKEDF
VEIARQKATVSHGATKGNDOVRELSCYLAPOIKVIAIRMPDEFYRNKGRNDL
MEYAKOHGIDIPVTPKRNPMMDENLMIHISEAGILENPKNOAPDGLYTTODPARAPN
TPDLIEIEFKGVVYVYKMDGTTHOTSELEMYENVEAGHVGGRIDIVENRIGM
KSRGIETTPAGTILYHAHLDIEAFITMDREVRIKQIGIKPAEIVYTGIRPSDEEFV
RHCIASQOEVEKGVQSVYKQGVYILGRESPLSLINBELVSMNVQGDYEPTDGTGI
NINSLRKEVHRIOKSVYTK"
variation         759..760
                  /note="pot. additional A in PAS1"
variation         1320
                  /note="U is C in variant PAS2"
variation         1431
                  /note="G is A in variant PAS1"
misc_feature      1526..1532
                  /note="put. polyadenylation signal"
misc_feature      1537..1544
                  /note="pot. polyadenylation signal"
polyA_site        1547
                  /note="polyadenylation site"
variation         1555
                  /note="A is U in variant PAS2"
BASE COUNT       393 a      431 c      434 g      302 t
ORIGIN
Query Match      100.0%; Score 1239; DB 9; Length 1560;
Best Local Similarity 100.0%; Pred. No. 4,1e-247;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 541 atgataagaacacctatcacatcagctacgagcttgaatccttgagaacccccagaac 600
DB 616 ATGGATGAGAACCTCATCATCAGCTACGAGGCTGGAAATCCTGGAGAACCCCAAGAAC 675
OY 601 caagcgcttcaggcttctacacgaagaccgaagaccgaagcccccaacacccct 660
DB 676 CAAGCGCTCCAGGCTCTACACGAGACCCAGACCACCAAGCCCCCAACACCCCT 735
OY 661 gacattctcagatcagatcctaaaggggtccctgtgaaggtagcaagctcaagat 720
DB 736 GACATCTCGAGATCGAGTTCAAAAAAGGGGTCCTGTGAAGGTGACCAACGTCAGAGAT 795
OY 721 ggcacaccccccaagacctcccttgagctcttcatlacttgaagaaagtgcgggcaag 780
DB 796 GGCAACACCCACGAGCTCTTGAGCTTTATGTACTGACGAGAGTCCGGGCAAG 855
OY 781 catggcttgaggcgtatctgaacatcgttgaagaccgcttcaatggaatgaagtcaggagt 840
DB 856 CATGGCGTGGGCGCTTATTGACATCGAGAGAACCGCTTCATTGGAATGAAGTCCGAGGT 915
OY 841 atcaagagaccccaagagccaccatccttaccatgctcatttgacatcgaagccttc 900
DB 916 ATCTACGAGACCCCGAGCAGCAGCACCATTCTTTACCATGCTCATTTGACATCGAGGCTTC 975
OY 901 accatgaccgggaagtgcgcaaatcaaacgaagccttgagcttgaaatttgctgagctg 960
DB 976 ACCATGGACCCGGGAGAGTGCCCAAAATCAAAAGGCGCTGGCTGAATTTCGTGAGCTG 1035
OY 961 gttatccgggtttaagcgttcgctgagctggttgatattgttcgcaatcgcgaag 1020
DB 1036 GTGATACCGGTTTACGGGCTTACGCTGCTGAGTGTGAATTTGTCGACATGCTGCCAAG 1095
OY 1021 tcccaagagcagctggaaggaagtgacagctgctgcctcctcaagggcgaagtgatc 1080
DB 1096 TCCAGAGAGGAGATGGAAGGAAAGTGAGGTGCTGCTCAAGGCGCAGCTGATCATC 1155
OY 1081 ctggcccggaaggtccccaatgctctctcacaatgagagctgtgtgaagacagctgcaag 1140
DB 1156 CTGGCGCGGAGTCCCACTGTCTCTCAATGAGGAGCGTGAGCATGAAGCTGCAG 1215
OY 1141 gttatattgaagcaactatccaccgggttaccatcaatcttccttaagctgaag 1200
DB 1216 GTGATTATGAGCAACTGATGCCACCGGTTTCATCAATCAATTCCTTCAGGCTGAAG 1275
OY 1201 gaatcatcgtctccagaagagctcaatgccaatag 1239
DB 1276 GAATATCATGCTCTCAGAGCAAGGTCACTGCCAATATG 1314

RESULT 2
AK027126      1975 bp      mRNA      PRI      29-SEP-2000
LOCUS        Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
DEFINITION   Homo sapiens cDNA: FLJ23473 fis, clone HS113532, highly similar to
ACCESSION   AK027126
VERSION      AK027126.1 GI:10440175
KEYWORDS     oligo capping; fis (full insert sequence).
SOURCE       Homo sapiens human small intestine cDNA to mRNA, clone_11b:HSI
              clone:HS113532.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE    1 (sites)
AUTHORS      Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
              Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
              Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.,
              NEDO human cDNA sequencing project
              Unpublished (2000)
JOURNAL      2 (bases 1 to 1975)
REFERENCE    2
AUTHORS      Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
              Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE        Direct Submission

```

JOURNAL

Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Submitted by: SUGANO, Institute of Medical Science, University of Tokyo, Sugenjo, Institute of Genome Structure Analysis, Human Genome Center, Laboratory of Genome Structure Analysis, Tokyo 108-8633, Japan. Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8633, Japan. shirokane-dai.4-6-1@u-tokyo.ac.jp, Tel:81-3-5449-5286, shirokane-dai.4-6-1@u-tokyo.ac.jp

(E-mail: [chana@nsl.ac.jp](mailto:chana@nsl.ac.jp))  
Fax: 81-3-5449-5416)  
NEO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency). Location/Qualifiers

```

FEATURES
source
location=/
1. .1975
/organism="Homo sapiens"
ccae"

```

```
misc_feature
1..1975
/note="highly similar to HSASD Human mRNA for
the poliovirus vaccine synthetase"
```

BASE COUNT	459	a	204	c	1
ORIGIN					

	99.54;	Score 1232.6;	DB 9;	Length 1975;
Query Match				
Best Local Similarity	99.77;	Pred. No. 8.4e-246;		
Matches 1235; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY           1 atgttcagcaagaagcccgcccgcgggcttgcacatcgtgatc 54  
              |||||  
Db       482 ATGTCAGCAAGCGCTCCGTGTTCTGGCTACAGTGCGCGCCTGCACACCTGTCATC 120

Dy 61 ctcgtatgctgaagaaacaaagcgcggcgccgcaacattggccag 60  
|||  
QY |||  
Db 542 CTCGTCGGCTGAAGAACAGCCTATGACGTCAATTGGCCAACTTGGCCAG 18

[illegible]

181 attcaggaatcagcaggagtcttgaggaggttccacggg  
|||||  
|||  
QY |||  
ATTTGAGCATGTCCAGCAGGAGTCTTGGAAGAGATTCACTGGCGGCCCATCCAGTCACGC

[illegible]

301 aaacaagtgnaatcgccacgcgaggcggaagccaaatgatgtgtccacacgagcgccacacgga  
|||||  
|||||ttcccccacgggagggggccacagatgtgtccacacgagcgccacacgga 8

[illegible]

D<sub>b</sub>      842 AAAGGAAACCAATCTTTTGGCCGTGCCTCCTTCATTGCGAATAATGACCTG  
QY    421 gtcattgtccctgagagtgcctaattctacaacggutcaaggcgcgaatgacctg

DQ 902 GTCAATGTCCTCCACAGAACCGTGGAGC  
db 481 atggaatgcgaagcaacacggattcccatcgcggtacttcccagaaccgctggagc

Dd 962 ATGGAGTACGCCAAGGCAATCACCGCCGCTGTTTCCTGAGAACCCCAAGAACC  
Oy 541 atggtatggaaacctcatctgcacatcagtcagtagtgtaattccctgagaaccaccaagaac

db 1022 ATGGATGAGAACCTCATGCACAC1CACC...  
ov 601 CAAGCGCCTCCAGTCTCTACAAGAGCCAGGCCAAGCCCCCAACCCCT

Db 1082 CAAGCCCTCCAGGCTCTACGACGAGACCCAGGCCAAGCCCCACACCCCT 1141  
 1142  
 1143  
 1144  
 1145  
 1146  
 1147  
 1148  
 1149  
 1150  
 1151  
 1152  
 1153  
 1154  
 1155  
 1156  
 1157  
 1158  
 1159  
 1160  
 1161  
 1162  
 1163  
 1164  
 1165  
 1166  
 1167  
 1168  
 1169  
 1170  
 1171  
 1172  
 1173  
 1174  
 1175  
 1176  
 1177  
 1178  
 1179  
 1180  
 1181  
 1182  
 1183  
 1184  
 1185  
 1186  
 1187  
 1188  
 1189  
 1190  
 1191  
 1192  
 1193  
 1194  
 1195  
 1196  
 1197  
 1198  
 1199  
 1200  
 1201  
 1202  
 1203  
 1204  
 1205  
 1206  
 1207  
 1208  
 1209  
 1210  
 1211  
 1212  
 1213  
 1214  
 1215  
 1216  
 1217  
 1218  
 1219  
 1220  
 1221  
 1222  
 1223  
 1224  
 1225  
 1226  
 1227  
 1228  
 1229  
 1230  
 1231  
 1232  
 1233  
 1234  
 1235  
 1236  
 1237  
 1238  
 1239  
 1240  
 1241  
 1242  
 1243  
 1244  
 1245  
 1246  
 1247  
 1248  
 1249  
 1250  
 1251  
 1252  
 1253  
 1254  
 1255  
 1256  
 1257  
 1258  
 1259  
 1260  
 1261  
 1262  
 1263  
 1264  
 1265  
 1266  
 1267  
 1268  
 1269  
 1270  
 1271  
 1272  
 1273  
 1274  
 1275  
 1276  
 1277  
 1278  
 1279  
 1280  
 1281  
 1282  
 1283  
 1284  
 1285  
 1286  
 1287  
 1288  
 1289  
 1290  
 1291  
 1292  
 1293  
 1294  
 1295  
 1296  
 1297  
 1298  
 1299  
 1300  
 1301  
 1302  
 1303  
 1304  
 1305  
 1306  
 1307  
 1308  
 1309  
 1310  
 1311  
 1312  
 1313  
 1314  
 1315  
 1316  
 1317  
 1318  
 1319  
 1320  
 1321  
 1322  
 1323  
 1324  
 1325  
 1326  
 1327  
 1328  
 1329  
 1330  
 1331  
 1332  
 1333  
 1334  
 1335  
 1336  
 1337  
 1338  
 1339  
 1340  
 1341  
 1342  
 1343  
 1344  
 1345  
 1346  
 1347  
 1348  
 1349  
 1350  
 1351  
 1352  
 1353  
 1354  
 1355  
 1356  
 1357  
 1358  
 1359  
 1360  
 1361  
 1362  
 1363  
 1364  
 1365  
 1366  
 1367  
 1368  
 1369  
 1370  
 1371  
 1372  
 1373  
 1374  
 1375  
 1376  
 1377  
 1378  
 1379  
 1380  
 1381  
 1382  
 1383  
 1384  
 1385  
 1386  
 1387  
 1388  
 1389  
 1390  
 1391  
 1392  
 1393  
 1394  
 1395  
 1396  
 1397  
 1398  
 1399  
 1400  
 1401  
 1402  
 1403  
 1404  
 1405  
 1406  
 1407  
 1408  
 1409  
 1410  
 1411  
 1412  
 1413  
 1414  
 1415  
 1416  
 1417  
 1418  
 1419  
 1420  
 1421  
 1422  
 1423  
 1424  
 1425  
 1426  
 1427  
 1428  
 1429  
 1430  
 1431  
 1432  
 1433  
 1434  
 1435  
 1436  
 1437  
 1438  
 1439  
 1440  
 1441  
 1442  
 1443  
 1444  
 1445  
 1446  
 1447  
 1448  
 1449  
 1450  
 1451  
 1452  
 1453  
 1454  
 1455  
 1456  
 1457  
 1458  
 1459  
 1460  
 1461  
 1462  
 1463  
 1464  
 1465  
 1466  
 1467  
 1468  
 1469  
 1470  
 1471  
 1472  
 1473  
 1474  
 1475  
 1476  
 1477  
 1478  
 1479  
 1480  
 1481  
 1482  
 1483  
 1484  
 1485  
 1486  
 1487  
 1488  
 1489  
 1490  
 1491  
 1492  
 1493  
 1494  
 1495  
 1496  
 1497  
 1498  
 1499  
 1500  
 1501  
 1502  
 1503  
 1504  
 1505  
 1506  
 1507  
 1508  
 1509  
 1510  
 1511  
 1512  
 1513  
 1514  
 1515  
 1516  
 1517  
 1518  
 1519  
 1520  
 1521  
 1522  
 1523  
 1524  
 1525  
 1526  
 1527  
 1528  
 1529  
 1530  
 1531  
 1532  
 1533  
 1534  
 1535  
 1536  
 1537  
 1538  
 1539  
 1540  
 1541  
 1542  
 1543  
 1544  
 1545  
 1546  
 1547  
 1548  
 1549  
 1550  
 1551  
 1552  
 1553  
 1554  
 1555  
 1556  
 1557  
 1558  
 1559  
 1560  
 1561  
 1562  
 1563  
 1564  
 1565  
 1566  
 1567  
 1568  
 1569  
 1570  
 1571  
 1572  
 1573  
 1574  
 1575  
 1576  
 1577  
 1578  
 1579  
 1580  
 1581  
 1582  
 1583  
 1584  
 1585  
 1586  
 1587  
 1588  
 1589  
 1590  
 1591  
 15

D6 1142 GACATTTCTCGAGTCGAGTTCATAAAAGGGGTCCCTCTGAAGTGACCACGTCAAAGAT 780

```
DQ      721 ggcacaccaccaccagaccccccccccggcgaggcgaag 126  
       |||||||  
DY      1202 ggcaaccacccaccagacctcttgaggacttcattgtacctgaaacaagtgcggggccaag 840
```

OY 781 catgagcttggtccgatcatgcacacgcgg-  
|||||  
1262 CATGGCGTGGCCGTATTGACATCTGGAGAACCCGCTTCATTGGAATCAAGTCCCGAGGT 13

841 atctacgagaccgccagcagcgacctcccccgcggcggttc  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
1322 attttagcagaccgccagcagcaccattcctttacatgcctcatcttagacatcgaggccttc  
- -

901 accatgacccgggaagtcgcgcaaatcaacaaggcccgaggcgtg  
|||||  
|||tttgcacccgggaagtccgcaaatcaaacaaagccctggccttgcaattttgctgaactg 14

[illegible][illegible][illegible]

Db 1562 CTGGGCCGGAGTCCGCTGTAAG  
QY 1141 ggtgattatgagccaactgtagtcacacggttcacatcaatlcacctagagctgtaag

Db	1622	GGTCATTATGAGCCAACTGATGACCTG	1239
Ov	1201	gaatatactctctccagagcaagtcactgccaatag	1720

Db 1682 GATATCATCGTCTCCAGACAAAGTCTC

RESULT	3	PRI	12-JUL-2
BC009243			
LOCUS	1595 bp	mRNA	
	BC009243		
	Homo sapiens, argininosuccinate synthetase, clone		MGC:3634

IMAGE:3010137, material  
BC009243 GI:14328058  
BC009243.1  
VERSION

SOURCE	ORGANISM	
human.	Homo sapiens	Cranialata; Vertebrata; Eutele
Eukaryota;	Metazoa;	Chordata;
Mammalia;	Primates;	Catarrhini; Homnidae; Homo

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 1595)	Strausberg, R.	Direct Submission
		National Institutes of Health, Ma
		Direct Submission
		National Institutes of Health, Ma

JOURNAL  
Submitted to: National Cancer Institute  
Gene Collection (MGC), Cancer Genomics  
Institute, 31 Center Drive, Bethesda, MD 20892  
USA  
http://mgc.nci.nih.gov

REMARK	COMMENT
NIH-MGC project	
Contact: MGC help desk	
Email: c9phbs-r@mail.nih.gov	
Procurement: ATCC	
Public laboratory	

CDNA Library prepared by: The I.M.A.G.E. Consortium  
CDNA Library Arrayed by: The Institute for Systems Biology  
DNA Sequencing by: [www.isb-sci.org](http://www.isb-sci.org)

http://www.systemsbio.org  
contact: amadan@systemsbio.org  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Fo

USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
Contact: [madan@systemsbiology.org](mailto:madan@systemsbiology.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Jullian

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://Image.llnl.gov/Sequence/Sequence.html>. This clone was selected for full length sequencing because it passed the following select:

**Source**

Location/Qualifiers: matched mRNA gi: 4557336.  
1. .1595

CDS

```

/db_xref="Homo sapiens"
/db_xref="LocusID:445"
/db_xref="taxon:9606"
/clone="MGC:3634 IMAGE:3010137"
/tissue_type="Muscle, Rhabdomyosarcoma"
/clone_lib="NH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
11.1319

```

BASE COUNT  
ORIGIN

400 a 443 c 440 g 312 t  
NINSLRKEVYHRLQSKVTAK"

Query Match	99.48;	Score 1231;	DB 9;	Length 1595;
Best Local Similarity	99.6%;	Pred. No. 1	9e-345.	
Matches 1234;	Conservative			

		0: Indels	0: Gaps	0:
QY	1 atgccacgaagaagctccgtagtctctgcccacacgtgaggcccttggaacctcgtgac	0:		
Db	81 ATGCCACGCAAAAGCTCCGCTGTTGGCCCTACAGTGGGGCCGTGACCCGTCATC		60	
QY	61 ctgcgtgctgacgaaagaacagagctatgacgctatgcctatctgccaacatctggccag		140	
Db	141 CTGCTGTGGCTCAAGAACAGAGCTTATGACGTCATTTGCTATCTGGCCAAATTGGCCAG		200	
QY	121 aaggaagacttcgaggaagccaggaagaagcagcgaagcttgggcccacaaagctgttc		180	
Db	201 AAGGAAGACTTCGAGGAAGCCACGAGAGAGGACATGAAGCTTGGGGCCAAAAGTGTTTC		260	
QY	181 attaggaatctcagcaaggagattctgtgaggaagctcatctggccggccatccacgtccacg		240	
Db	261 ATTAGGATGTCACGACGAGGATTGTGGAGGAGTTCATCTGGCCGGCCATCCACGTCCACGC		320	
QY	241 gcactgtatgaggaacgcgtacctctctgggaacctctctggcaggacctgcatctggccgcg		300	
Db	321 GCACCTGATGAGAGACCGCTACCTCCTGGGACCTCTTGTGCCAGGACCTGTGATGGCCGC		380	
QY	301 aaacaagctggaatctgccagcgggaaggggccaaatagtgctccacgagccacagga		360	
Db	381 AAAACAGTGGAAATCGCCACGCGAGGGGCCAAGTGTGTCCACAGGCGCCACAGGA		440	
QY	361 aaggggaaagcatcagtcgcggtttagctcagctctactactcagggccccccagataaag		420	
Db	441 AAGGGGAACGATCAGTCCGCTTTAGGCTCAAGCTGTACTCATCGCCCCCCACGAATTAAG		500	
QY	421 gtcatctgcctcccgaggatgcctgaaatctcaaacacggttaaggccgaatctaacgt		480	
Db	501 GTCATCTGCTCCCTCGAGGAGTGCCTGAATTTACACCGGTTCAAGGGCCGCAATGACCTG		560	
QY	481 attaggaagcaagaacacagggattccatccacccggttaactccccaagaacccgtggagc		540	
Db	561 ATTAGGATGACCAAAACACACAGGGATTCCCATCCGCTACTCCCAAGAACCCGTGGAGC		620	

[illegible]

RESULT	4
AC027153	
LOCUS	
DEFINITION	AC027153 180838 bp DNA
ACCESSION	Homo sapiens chromosome 17 clone RP11-54I03 HTG
VERSION	AC027153.1
KEYWORDS	SEQUENCE, 28 unrelated pieces.
SOURCE	AC027153.1 GI:7731447
ORGANISM	HTG: HTGS_PHASE1; HTGS_DRAFT. human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 180838) Birren,B., Linton,L., Nusbaum,C. and Landier,E.
JOURNAL	Homo sapiens chromosome 17, map 17, WORKING DRAFT
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 180838) Birren,B., Linton,L., Nusbaum,C. and Landier,E.

Anderson, S., Baskin, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Boguslavsky, L., Baldwin, J., Barra, N., Bastien, V., Beda, F., Campopiano, A., Bokkhaaler, B., Brown, A., Burkett, G., Collymore, A., Castle, A., Choegel, Y., Colangelo, M., Collins, S., Dodge, S., Domino, M., Doyle, M., Dearellano, K., Dewar, K., Diaz, J., Dodgson, P., Domin, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

TITLE  
JOURNAL  
COMMENT

Galagan, J., Gardyna, S., Gande, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Lakoque, K., Lamazares, R., Landers, J., Lenocksky, J.,  
Levine, R., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
McCarthy, M., McMan, P., McGurk, A., McKernan, K., McPheters, R.,  
Meidirm, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information

Center project name: L5602

Center clone name: 541\_O\_3

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 169058 bases at least Q40

Consensus quality: 174233 bases at least Q30

Consensus quality: 174233 bases at least Q20

Insert size: 179000; agarose-fp

Insert size: 178138; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

consists of 28 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 315: contig of 315 bp in length

316 415: gap of 100 bp

416 1503: contig of 1088 bp in length

1504 1603: gap of 100 bp

1604 3415: contig of 1812 bp in length

3416 3515: gap of 100 bp

3516 5368: contig of 1833 bp in length

5369 5468: gap of 100 bp

5469 7164: contig of 1696 bp in length

7165 7264: gap of 100 bp

7265 9123: contig of 1859 bp in length

9124 9223: gap of 100 bp

9224 10641: contig of 1418 bp in length

10642 10741: gap of 100 bp

10742 11946: contig of 1205 bp in length

11947 12046: gap of 100 bp

12047 14424: contig of 2378 bp in length

14425 14524: gap of 100 bp

14525 17721: contig of 3197 bp in length

17722 17821: gap of 100 bp

17822 20638: contig of 2817 bp in length

20639 20738: gap of 100 bp

20739 24567: contig of 3829 bp in length

24568 24667: gap of 100 bp

FEATURES  
source

24668 29147: contig of 4480 bp in length  
29148 29247: gap of 100 bp  
29248 33430: contig of 4183 bp in length  
33431 33530: gap of 100 bp  
33531 38145: contig of 4615 bp in length  
38146 38245: gap of 100 bp  
38246 43517: contig of 5272 bp in length  
43518 43617: gap of 100 bp  
43618 48556: contig of 4939 bp in length  
48557 48656: gap of 100 bp  
48657 55104: contig of 6448 bp in length  
55105 55204: gap of 100 bp  
55205 55833: contig of 4629 bp in length  
55834 59933: gap of 100 bp  
59934 66011: contig of 6078 bp in length  
66012 66111: gap of 100 bp  
66112 72673: contig of 6562 bp in length  
72674 72773: gap of 100 bp  
72774 79925: contig of 7157 bp in length  
79926 80025: gap of 100 bp  
80026 85367: contig of 5342 bp in length  
85368 85467: gap of 100 bp  
85468 92616: contig of 7149 bp in length  
92617 92716: gap of 100 bp  
92717 101894: contig of 9178 bp in length  
101895 101994: gap of 100 bp  
101995 118357: contig of 16363 bp in length  
118358 118457: gap of 100 bp  
118458 138888: contig of 20431 bp in length  
138889 138988: gap of 100 bp  
138989 180838: contig of 41850 bp in length.

Location/Qualifiers

1. 180838

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="17"

/map="17"

/clone="RP11-54103"

/clone\_lib="RPCT-11 Human Male BAC"

1. 315

/note="assembly-fragment"

clone\_end:17

vector\_side:left

416 1503

/note="assembly-fragment"

1604 3415

/note="assembly-fragment"

3516 5368

/note="assembly-fragment"

5469 7164

/note="assembly-fragment"

7265 9123

/note="assembly-fragment"

9224 10641

/note="assembly-fragment"

10742 11946

/note="assembly-fragment"

clone\_end:sp6

vector\_side:right

12047 14424

/note="assembly-fragment"

14525 17721

/note="assembly-fragment"

17822 20638

/note="assembly-fragment"

20739 24567

/note="assembly-fragment"

24668 29147

/note="assembly-fragment"

29248 33430

/note="assembly-fragment"

33531 38145

/note="assembly-fragment"

misc_feature	38246.	43517	/note="assembly-fragment"
misc_feature	43618.	.48536	/note="assembly-fragment"
misc_feature	48657.	.55104	/note="assembly-fragment"
misc_feature	55205.	.59833	/note="assembly-fragment"
misc_feature	59934.	.66011	/note="assembly-fragment"
misc_feature	66112.	.72673	/note="assembly-fragment"
misc_feature	72774.	.79925	/note="assembly-fragment"
misc_feature	80026.	.85367	/note="assembly-fragment"
misc_feature	85468.	.92616	/note="assembly-fragment"
misc_feature	92717.	.101894	/note="assembly-fragment"
misc_feature	101995.	.118357	/note="assembly-fragment"
misc_feature	118456.	.118888	/note="assembly-fragment"
misc_feature	138989.	.180838	/note="assembly-fragment"
misc_feature	180838.	.54235	/note="assembly-fragment"
misc_feature	54235	2701	others

Query Match	88.2%;	Score 1093;	DB 2;	Length 180838;
Best Local Similarity	93.7%;	Pred. No. 3.4e-217;		
Matches 1161;	Conservative	0;	Mismatches 75;	Indels 3;
			Gaps	2;

QY	1	atgccaagcaagagctcgttggtctctggcctacagttgcygacctgagaccctgtgcatc	60
Db	147133	ATGTCCACCAAAAGGCTCCCTGGTTCTGGCTACAGTGGCACCTGGACACTCTCTGGACC	147192
QY	61	ctcgtgtggtctaagaanaacagcgtacagctcatcttgccatctctggccaacttggccag	120
Db	147193	CTCTGTGGCTGAAGGACMACAGGCTATATAGTGCTATCGCTACCTGGCAACATTTGGCCAG	147252
QY	121	aaggaagacttcgaggaagcccgagaaagacatlgagactltggtggccaaagaagtgctc	180
Db	147253	AAGGAAGACTTCGAGGAAGCCGAGAAAGTAATGACTGAAAGTTTGGGGCTAAAGAAGCTTTC	147312
QY	181	atggaagatgtaagaaggaaggtttgttggaagatcactctgycgcgcacatccagtcacgc	240
Db	147313	ATTGAGAGATGTGGCAGGAGATGTTGTGGAGAGATCTATCTGGCGCGGCATCCAGTCCAGT	147372
QY	241	gcacttctagagagaccgtacctccctggacacctctcttcagagccctgaaatgscgcgc	300
Db	147373	GCACCTATGAGGACCGGTACCTTCTGGGCACCTCTC--GCCAGGCCCTGCACTCGCTCC	147430
QY	301	aaacaagtggaaatcgcgccagcggaggaaggggccaaglatgtgtcccaagcgcgcacagga	360
Db	147431	AAACAAMTGAATATCCCACTGGAAAGGGGGCCAAAGTATATGTCCACAGCACCAAGGGA	147490
QY	361	aaggggaagcatcaggtccggtttggaactaagctgctactcactctggtgccccagataaag	420
Db	147491	AATGGGCAACATCCAGTCTCGGTGTGAGCTCATTTGGTACTACCTATACCCCCAGATAAAG	147550
QY	421	gtcaattgctcccttggaaatgctcgaaattctacaacccggttcaagggccgaatgacctg	480
Db	147551	GTCATTGTCCTCGTGGAGATGCCCAATTTCTACACCAAGCTCAAGGGCTCTCATATACCTG	147610
QY	481	atggaggtctgcgaagcaaacacgggaattcccatcccggtcaactcccaagaacccttggagc	540
Db	147611	ATGGAGATATGGAAAGCAACAGAGATTCCTCATCCGGTCAACACCAAAGAACTCGTGGAGC	147670
QY	541	atggaatggaacctctgacatatagcttagaagctggaatctctgtgagaaacccaagaac	600
Db	147671	ATGGACGAGAACTCTATGCTATATAGCTACGAGTCTGGGAATCTTGGAAGCCCTTAAGAAC	147730

QY	601	caaggccctccagagctctcaagcgaagcccgagcccaagcccaagcccccaacccct	660
Db	147731	CAAGCACTCCAGGCTCTACACGAAGA	CCAGGACCGGGCCAAAGCCCCCAACACCCCT
			147788
QY	661	gacatctcagagatccaggttcaaaaaaggggtccctgtgaaggtgaaccaagtcgaagat	720
Db	147790	GACATTCTTGAGATGAGATTCCAAAAAGAGTCCCTGTGAAGGTGACCAACGTCAGGAT	147845
QY	721	ggcacaccacacagaaccctcttgagctcttcaatgtaacctggaacgaagtcggggcaag	780
Db	147850	GGCACACCACCAACGACGTCATTTGAGGCTTTCATGTACTGAACGAAGTGTGGGCAAG	147905
QY	781	catgctgtagggcgatattgacatcgltgagaacgctcattcgtgaatgaagtcccagagt	840
Db	147910	CAGGGTGTGGGCTGTATTGACTCATGAGAAACGGCTTCATTGGAATGAAGTCCCGAGGG	147965
QY	841	alctacggaaccccaagcagacacatcccttacaatgctcatttagaactgaagccttc	900
Db	147970	ATCTACCGAGGCCCCAGCAGACACCACTCTTACCTTGACCTTTAGACACCAAGGCTTTC	148029
QY	901	accatggagccgggaagtcgcgaataatcaaaaaagcctggagcttgaatttgcctgagctg	960
Db	148030	ACCATGGGCCAGAGAGTGTCCGAATTCAAACAAGGCGCTTGAGAAATTTGTTGAGCTTG	148089
QY	961	gtgtataccggcttaacggctagccctgaagtgtgaatttgcgcgaactgcattcgccaag	1020
Db	148090	GTGGATTACCGGTTTCTGGTACAGCCCTGAGCTGGAATTTGCTGCGCCACTGCATCGCCAG	148149
QY	1021	tcccagagcgagatggaaaggaagatgcaggtgcgctctccaagggccagagtgataatc	1080
Db	148150	TCCAGGAGCAAGTGGAAAGGAAGTCCAGGTGTCCGTTCTTCAGAGGGCCAGGTTGTCATTC	148209
QY	1081	ctcgagccggaggtccccaactgtctcttacaatgaagagcttgltgagcaatgaactgacag	1140
Db	148210	CTTGCCCAAGGAGTCCCACTGTCTCTCTACAAAGAGAGAGCGCGGTGAGCATGAACGTGCAG	148269
QY	1141	gggtattatgagccaactgtatgcacacgggttcatacaactcaatctccctcagctgaag	1200
Db	148270	GGGTATTATCAGCAATTGATGGCCACCGGGTTTATCAACTCAATTCCTCTCAGGCTGAAG	148329
QY	1201	gaatatcatgctctccaggaagaagtgtaactgcgcaaatag	1239
Db	148330	GAAATATCATGCTCTCCAGACGAAGGTGACGTACGCCCAATATG	148368

```

RESULT      5
AC025175/c
LOCUS
DEFINITION  AC025175      165439 bp      DNA      HTG      18-JUL-2000
AC025175      Homo sapiens chromosome 5 clone CTD-208J1C10, WORKING DRAFT
AC025175      SEQUENCE, 20 ordered pieces.
AC025175
AC025175.3      GI:9256455
HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
ORGANISM    human.
             Homo sapiens
Makayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165439)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 165439)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7711790.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----

```









AC009550	LOCUS	DEFINITION	AC009550	188356 bp	DNA	HTG	20-APR-2000
AC009550	Homo sapiens clone RP11-382M14, WORKING DRAFT SEQUENCE, 11 unordered pieces.						
AC009550	AC009550.3	GI:7622313					
AC009550	HTG: HTGS_PHASE1; HTGS_DRAFT.						
AC009550	Homo sapiens chromosome, clone RP11-382M14						
AC009550	Unpublished						
AC009550	2 (bases 1 to 188356)						
AC009550	Blumenfeld, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Baran, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Deary, E., Devon, K., Dewar, K., Donnell, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Garayna, S., Gilbert, D., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J.C., Jones, C., Kann, L., Karas, A., Lebeck, J., Lien, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K., McLaughlin, J., Melnick, J., Moll, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Senger, R., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Toriella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.						
AC009550	Submitted						
AC009550	Submitted (27-AUG-1999)						
AC009550	Research, 320 Charles Street, Cambridge, MA 02141, USA						
AC009550	On Apr 20, 2000 this sequence version replaced gi:6006236.						
AC009550	All repeats were identified using RepeatMasker:						
AC009550	Smith, A.F.A. & Green, P. (1996-1997)						
AC009550	http://ftp.genome.washington.edu/RM/RepeatMasker.html						
AC009550	Genome Center						
AC009550	Center: Whitehead Institute/ MIT Center for Genome Research						
AC009550	Center code: MIBR						
AC009550	Web site: http://www.seq.wi.mit.edu						
AC009550	Contact: sequence_submissions@genome.wi.mit.edu						
AC009550	Project Information						
AC009550	Center project name: L1662						
AC009550	Center clone name: 382.M14						
AC009550	Summary Statistics						
AC009550	Sequencing vector: M13, M7815; 100% of reads						
AC009550	Chemistry: Dye-terminator; 4% of reads						
AC009550	Chemistry: Dye-terminator; 96% of reads						
AC009550	Assembly program: Phrap; version 0.960731						
AC009550	Consensus quality: 159465 bases at least Q40						
AC009550	Consensus quality: 175637 bases at least Q30						
AC009550	Consensus quality: 182085 bases at least Q20						
AC009550	Insert size: 194000; agarose-fp						
AC009550	Insert size: 187356; sum-of-contigs						
AC009550	Quality coverage: 4.6 in Q20 bases; agarose-fp						
AC009550	Quality coverage: 4.6 in Q20 bas.						
AC009550	NOTE: This is a 'working draft' sequence. It currently						
AC009550	* consists of 11 contigs. The true order of the pieces						
AC009550	* is not known and their order in this sequence record is						
AC009550	* arbitrary. Gaps between the contigs are represented as						
AC009550	* runs of N, but the exact sizes of the gaps are unknown.						
AC009550	* This record will be updated with the finished sequence						
AC009550	* as soon as it is available and the accession number will						
AC009550	* be preserved.						
AC009550	1						
AC009550	1599 1798: contig of 1698 bp in length						
AC009550	1799 4670: contig of 2872 bp in length						
AC009550	4671 4770: gap of 100 bp						
AC009550	4771 8967: contig of 4197 bp in length						
AC009550	8968 9067: gap of 100 bp						
AC009550	9068 13012: contig of 3945 bp in length						

Query Match	Best Local Similarity	Matches 1151: Conservative	0: Mismatches	85: Indels	12: Gaps	3: Others
1	atgccgcaagaagcgcgcgltgtctgagcctacaatgacgagccgctgagacactcgtgtgac	60				
Db 121401	ATGCCAGCAAGAGGCTCTATGTTCTGCGCTGCACTGACCTGGACCTGGACACCTTCTGCTAT	121460				
Qy 61	ctcgtgtgctgaagaaacaagcgtatgacgtcattgtgctactatctgagcaaatgagccag	120				
Db 121461	CTCGTGTGCTGAAGAAACAAGGCTATGACGTCATTGCTCTACCTGAGGCGCAGATTGGCCAG	121520				
Qy 121	aaggaagactctgaggaagcccgaggaagagcgtgaagccttgggtgcccagaaggtgttc	180				
Db 121521	AAGGAAGACTTCCAGGAAGCCAGGAAGAGCCTGAGGCTGGGGCCAAAAGATGTGA	121580				
Qy 181	atgagagatctcagcagagagcttctgagagatctcaactctgcccgcacacagctccag	240				
Db 121581	ATTGAGAGATGCTCAACAGGAGATTGTGTGAGGAGATTCACTGTGCGCGGCATGCACTCAGC	121640				
Qy 241	gcacctatgaggaacgcctactcctctggtgacactctcttgcagagccctgcatgcgcg	300				
Db 121641	ACACGTATATGAGGACCGCTACCTCTCTGGAGACCTCTGCGCAGGCGCCTCATCACCCAG	121700				



*	31507	31606:	gap of	100 bp
*	31607	63899:	contig of 32284 bp	in length
*	63891	63990:	gap of	100 bp
*	63991	91527:	contig of 27537 bp	in length
*	91528	91637:	gap of	100 bp
*	91628	106175:	contig of 14548 bp	in length
*	106176	106275:	gap of	100 bp
*	106276	115509:	contig of 9234 bp	in length
*	115510	115609:	gap of	100 bp
*	115610	125111:	contig of 9502 bp	in length
*	125112	125211:	gap of	100 bp
*	125112	133211:	contig of 8000 bp	in length
*	133212	133311:	gap of	100 bp
*	133312	142573:	contig of 9262 bp	in length
*	142574	142673:	gap of	100 bp
*	142674	149569:	contig of 6886 bp	in length
*	149570	149669:	gap of	100 bp
*	149670	155266:	contig of 5597 bp	in length
*	155267	155366:	gap of	100 bp
*	155367	160596:	contig of 5230 bp	in length
*	160597	160696:	gap of	100 bp
*	160697	165212:	contig of 4516 bp	in length
*	165213	165312:	gap of	100 bp
*	165313	171081:	contig of 5769 bp	in length
*	171082	171181:	gap of	100 bp
*	171182	174820:	contig of 3639 bp	in length
*	174821	174920:	gap of	100 bp
*	174921	178093:	contig of 3173 bp	in length
*	178094	178193:	gap of	100 bp
*	178194	182424:	contig of 4231 bp	in length
*	182425	182524:	gap of	100 bp
*	182525	183733:	contig of 1209 bp	in length
*	183734	183833:	gap of	100 bp
*	183834	186908:	contig of 3075 bp	in length
*	186909	187008:	gap of	100 bp
*	187009	189277:	contig of 2269 bp	in length
*	189278	189377:	gap of	100 bp
*	189378	193066:	contig of 1288 bp	in length

[illegible]

```

Db 50462 GCACAGCGCTGGGCAATATTCGATCGTGAGACACCGTTTCATGGAAGAGCCGAGC 50403
Oy 840 tatctagaagaccagagagacacatctcttaccagtcatttaacacagagccct 899
Db 50402 TATCTACCGGGCCCGACAGGACACCATCTTGTACGCTATTAGCATCGAGGCGCT 50343
Oy 900 caccatgagccgggaagagcgcaaatcaaacagagccctgagctgaattgctgagct 959
Db 50342 CACCTGTACCGGAGAGGCAACAAATCAATGCGGGGCTTGAAATTTGCTGAGCT 50283
Oy 960 ggtgtataccggttaagcgtct-----agccctgagtgtaattgttcgcacatgc 1011
Db 50282 GGTGTATCTACTGTTCTGCTGCTTCCGGCAGACCCCTGAGTGTGATTTGTCGCCACTGCG 50223
Oy 1012 atgcgcaagctcccgagagcgagtggaagggaagtgacgagtgctcgtctcaagggcag 1071
Db 50222 ATCGCCGAGTCCAGAGATGAGTGAGGGAAGTGCGAGTGCTGCTTAAAGGCGCAG 50163
Oy 1072 ggtacacctcgcgcgagagatcccaactgctcctacaaatgagagcgctgtagatg 1131
Db 50162 GTGTACATCTTGGCTGGAGATGCCC---CACTGTCTTACACGAGAGCCGCTGAGCATG 50106
Oy 1132 aacgtgcagggatgattatgagccaactgagccacccgggttcacatcaatcaatccctc 1191
Db 50105 AACGTGACGAGGCTGATATGACCAATTCATCCACCGGCTTCATCAACATCAATTCCTCCTC 50046
Oy 1192 aggcgaaggaatcatcatcctcctcagagcaaggtcactgccaatag 1239
Db 50045 AGCGTGAAGGANTATCATCTGCTCCAGAGCAAGTCACTGCAAAATAG 49998

RESULT 9
AC073169 201989 bp DNA HMG 06-MAY-2001
LOCUS Homo sapiens chromosome 11 clone RP11-87E1 map 11, WORKING DRAFT
DEFINITION AC073169
AC073169.1 GI:8389521
VERSION HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 11, clone RP11-87E1
REFERENCE 2 (bases 1 to 201989)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bouslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeBreliano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labrecque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPherson,R.,
Meldrum,T., Meneses,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Theodore,J., Titrill,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

## COMMENT

All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 110271

Center clone name: 87\_E1

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 189608 bases at least Q40

Consensus quality: 198358 bases at least Q20

Insert size: 205000; agarose-fp

Insert size: 200189; sum-of-ctnigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 19 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

```

1 1275: contig of 1275 bp in length
1276 1375: gap of 100 bp
1376 2986: contig of 1611 bp in length
2987 3086: gap of 100 bp
3087 4668: contig of 1582 bp in length
4669 4768: gap of 100 bp
4769 7645: contig of 2877 bp in length
7646 12379: gap of 100 bp
7746 12379: contig of 4634 bp in length
12380 12479: gap of 100 bp
12480 16061: contig of 3582 bp in length
16062 16161: gap of 100 bp
16162 21161: contig of 5000 bp in length
21162 21261: gap of 100 bp
21262 26219: contig of 4958 bp in length
26220 26319: gap of 100 bp
26320 34818: contig of 8499 bp in length
34819 34918: gap of 100 bp
34919 40947: contig of 6029 bp in length
40948 41047: gap of 100 bp
41048 46162: contig of 5115 bp in length
46163 46262: gap of 100 bp
46263 56587: contig of 10325 bp in length
56588 56687: gap of 100 bp
56688 66612: contig of 9925 bp in length
66613 66712: gap of 100 bp
66713 78573: contig of 11861 bp in length
78574 78673: gap of 100 bp
78674 92045: contig of 13372 bp in length
92046 92145: gap of 100 bp
92146 109655: contig of 17510 bp in length
109656 109755: gap of 100 bp
109756 135749: contig of 25994 bp in length
135750 135849: gap of 100 bp
135850 164544: contig of 28695 bp in length
164545 164644: gap of 100 bp
164645 201989: contig of 37345 bp in length.

```

## FEATURES

source

1. 201989  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"

```

/map="11"
/clone="RP11-87E1"
/clone_lib="RPCT-11 Human Male BAC"
1..1275
misc_feature
  /note="assembly_fragment"
  1376..2286
misc_feature
  /note="assembly_fragment"
  3087..4668
misc_feature
  /note="assembly_fragment"
  4769..7645
misc_feature
  /note="assembly_fragment"
  7746..12379
misc_feature
  /note="assembly_fragment"
  12480..16061
misc_feature
  /note="assembly_fragment"
  16162..21161
misc_feature
  /note="assembly_fragment"
  21262..26219
misc_feature
  /note="assembly_fragment"
  26320..34818
misc_feature
  /note="assembly_fragment"
  34919..40947
misc_feature
  /note="assembly_fragment"
  clone_end:77
vector_side:right"
  41048..46162
  /note="assembly_fragment"
  clone_end:sp6
  vector_side:left"
  46263..56587
misc_feature
  /note="assembly_fragment"
  56688..66612
misc_feature
  /note="assembly_fragment"
  66713..78573
misc_feature
  /note="assembly_fragment"
  78674..92045
misc_feature
  /note="assembly_fragment"
  92146..109655
misc_feature
  /note="assembly_fragment"
  109756..135749
misc_feature
  /note="assembly_fragment"
  135850..164544
misc_feature
  /note="assembly_fragment"
  164645..201989
  /note="assembly_fragment"
BASE COUNT 65404 a 37765 c 36760 g 60255 t 1805 others
ORIGIN

```

```

Query Match      85.4%; Score 1058; DB 2; Length 201989;
Best Local Similarity 92.2%; Pred. No. 6e-210;
Matches 1151; Conservative 0; Mismatches 85; Indels 12; Gaps 3;

QY 1 atgtccagaagaagtcggtgtgttcgtgacctacagtgvgcgccctggaacctcgtgcatc 60
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196606 ATGTCCAGCAAAAGGCTCTAAGTGTGTGGCTTGCGAGTGGAGCCCTGGACACCTTCTGCATC 196665

QY 61 ctgctgtgtgtgtgaagaagaagctatgacgtcatctcctatctgccaacattggccag 120
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196666 CTGCTGTGTGTGTGAAGAACAGGCTATGACGTCTATCTCTACTCTGGCCAGATTGGCCAG 196725

QY 121 aaggaagaacttcgaagaagccgaagaagaagcactgaagcttggggccaaaggtgttc 180
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196726 AAGGAAGACTTCCAGAGAGCCAGAGAGACACTGAAGCCTGGGGCCAAAGATGTTTA 196785

QY 181 atgaagatgtcagcagaagagtttgtgagagatcatctgtgcccgcacatccagtcacg 240
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196786 ATTGAGAGTGTCAACAGGAGATTGTGTGAGAGATTCACTGTGGCCGACATGACATGCCAGC 196845

QY 241 gcactgtataggaagcactcctcctctggaacacctctctcccaagagccctgcatcgccgc 300
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196846 ACATGTTATGAGGACCGCTACTCTCTGGGAACCTCTCTCGCCAGAGCCCTGATATCACCAC 196905

```

```

QY 301 aaacaagtggaaatcgcaccagcgggaaggggccaaagtatgtgtcccaagcgccacaga 360
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196906 AAACAAGTGGAAATCGCCAGGAGGAGGGGCAAAATATGTGTCCAGGACCAACAGGAGA 196965

QY 361 aagggaagaagatcaggtcgtgtgtgagctcagctcctactcaactggtccccc-caagataa 419
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 196966 AAGGGGAATGATCAGGTCTGTGTGAGCTCACCTCTACTCTCTCTGCCCCCAGATATAA 197025

QY 420 gttcattgtcttccttgagagatgctgaattcttacaaccggttcaaggccgcaatgacct 479
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197026 GCTCATGTGCTCTCTGATGATGCCCAATTCCTACAAACCGGTTCAAGGGCCACAGTACCT 197085

QY 480 gatggagtcagcaagaacacacggaattcccaatcccggtctactcccaagaacctgtgag 539
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197086 GAGCGAATATGCAAAAGCAACGCGGAGATTCCACCCCGGTCACTCCCAAGAACCCCTTGAG 197145

QY 540 catgatatggaacctcatcatcatcagctcagctcagagcttgaaatccctggaagaccaccaaa 599
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197146 CATGGAAGAACTCTCATGACATCAGCTGCGAGGCTGGAATCTGGAGAAATCCCGAGAA 197205

QY 600 ccaagcgctccaggtctctctacacgaagaccacagaccgaagcccaaacacccc 659
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197206 CCAAGCGCTCTCAGCTCTCTACACGAAGCCAGGACCCGCGCAAAACCCCAACACCC 197265

QY 660 tgaactctcgaatcagatcgaatcaaaaagggttccctgtgaagtgtaccacagtcagaaga 719
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197266 TGACATTTCTCGAGATCGAATGCAAAAGGGGTCCCGGTAAGAGTCAACCAAGTCAAGGA 197325

QY 720 tggcaaccacccacagacctctctgtgagctctctatgtaactgaacgaagtgcggggcaa 779
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197326 TGATGCGACACCCACAGACCTCTTGAGCTCTTCAATGACCGAAGGAGTGGCGGGA 197385

QY 780 gcatgctggtgacgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 839
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197386 GCACAGCGTGGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 197445

QY 840 tatctagaagaccacagacagacacacacacacacacacacacacacacacacacacacac 899
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197446 TATCTACCGGGGCCCGCCAGCAGACCATCTTGTGATGATGATGATGATGATGATGATG 197505

QY 900 cccactgagccgggaagtgccgaataaagaagccctggaattgtgtgagct 959
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197506 CACCATGTACCGGGGAAGTCCACAAATAATCAACAAATGCCGGGCTTGAAATTTGCTGAGCT 197565

QY 960 ggtgtatcaggtttacggcct-----agccctgaagtgtgaattgttcggccatgctc 1011
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197566 GGTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 197625

QY 1012 atgcgaagtcaccagaagcagtggaaggaaggaagtgcaagtggtccgtcccaaggccag 1071
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197626 ATGCGCGAGTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 197685

QY 1072 ggtatattctcgtccgggaagtcaccacgtctctctaaatgaagaagctgtgtgagcag 1131
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197686 GGTGTATCTCTTGGCGGAGTGTCCC---CACTGTCTAAAGAGAGAGCGGTGTGAGCATG 197742

QY 1132 aacgtgcaagtgatattatagccaactgtatgcacacgggtttatcaatacaatcctcctc 1191
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197743 AAGCTGACAGGATATATATAGCAATATGATGATGATGATGATGATGATGATGATGATG 197802

QY 1192 aggtgaaagaalatalcatcgtctccagagcaaggtcaactgtccaaatag 1239
  ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 197803 AGGCTGAAGGAATATATCATGTCTCCAGACCAAGGTCACATGCAAAATAG 197850

RESULT 10
AL353717/c 155019 bp DNA HTG 05-AUG-2001
LOCUS Homo sapiens chromosome 9 clone RP11-562M8, *** SPOUNDCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL353717
VERSION AL353717.9 GI:15131977
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFLN; HTGS_DRAFT; HTGS_PULLTOP.

```

```

SOURCE          human.
ORGANISM        Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE           1 (bases 1 to 155019)
JOURNAL         Direct Submission
                Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
                requests: clonerequests@sanger.ac.uk
                On Aug 9, 2001 this sequence version replaced gi:15131481.
COMMENT         -----
                Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                -----
                Project Information
                Center project name: ba562m8
                -----
                Summary Statistics
                Assembly program: XGAP4, version 4.5
                Sequencing vector: M13; M77815; 0% of reads
                Sequencing vector: plasmid; L08752; 99% of reads
                Chemistry: Dye-terminator ET-amersham; 0% of reads
                Chemistry: Dye-terminator Big Dye; 99% of reads
                Chemistry: Dye-primer-amersham; 0% of reads
                Consensus quality: 154190 bases at least Q40
                Consensus quality: 154545 bases at least Q30
                Consensus quality: 154709 bases at least Q20
                Insert size: 154819; sum-of-contigs
                Insert size: 165328; 7.3% error; agarose-fp
                Quality coverage: 8.29x in Q20 bases; sum-of-contigs quality
                coverage: 7.89x in Q20 bases; agarose-fp
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 3 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                1 97615: contig of 97615 bp in length
                * 97616 97715: gap of 100 bp
                * 97716 117003: contig of 19288 bp in length
                * 117004 117103: gap of 100 bp
                * 117104 155019: contig of 37916 bp in length.
                Location/Qualifiers
                1.155019
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone="RP11-562M8"
                /clone.lib="RPC1-11.2"
                1.97615
                /note="assembly-fragment:03117
                fragment_chain:1"
                misc_feature
                97716..117003
                /note="assembly-fragment:01189
                fragment_chain:1"
                misc_feature
                117104..155019
                /note="assembly-fragment:02543
                fragment_chain:1"
                BASE COUNT 42308 a 31372 c 33444 g 47694 t 201 others
                ORIGIN
                Query Match 85.4%; Score 1057.8; DB 2; Length 155019;
                Best Local Similarity 92.5%; Pred. No. 6.9e-210;
                Matches 1146; Conservative 0; Mismatches 87; Indels 6; Gaps 3.

```

OY	61	ctctgtgctgtgaagaaacaaagcctctctgacgtcatctgctctctctgcccacattgtggcag	120
Db	65336	CTCTGTGTGCTGTAAGAACAAAGCCTTGTGACATCATTTGCTTACTCTGTGCGACACTGTGGCAG	6527
OY	121	aagaaagactctgaggaagaccgaagaaagagcactgaagctcttgggtccaaagatgtctc	180
Db	65276	AAGAAAGACTCTGAGAAAGCCTTGAAGAGGACGACTGAAGCTTTGGACCAAAAAGTGTTC	6521
OY	181	attgagatgtctagccagggagtttttggagagttcatctctgcccgcattccagtcacg	240
Db	65216	ATTGAGATGTCTAGCCAGGAGGAGTTTGTGGAGGATTCATCTGACCGCATTCACATCCAGC	6515
OY	241	gcactgtataggaaccgctactcctctggcaccctctctctgcagaacctgaatgcgccg	300
Db	65156	GCACGTATATGAAACCGCTACTCCCTGGGCACTCTCTGAGGACAGCCCGGATGTGCCGCG	6509
OY	301	aaacaagtgtgaatctgcgccagcgggaggggccaagtatgtctcccaagtcgcacaga	360
Db	65096	AAACAAGTGTGAATCTGCCAAGAGGAGGAGGCCAAGTATGTGTTCACAGGACACCGGGA	6503
OY	361	aaggaggacgatacaggtccggttttagcttcagcgtctcactcaactcgtgccccagataaag	420
Db	65036	AAAGGAATATATAGTTCGGGTTTGAGCTCACCTCTATTGTGCTGGCCCCCAGATTAAG	6497
OY	421	gtcatgtctccctgtagagatgctctgaattctcaaacccgttcaagggtccgcaatgtacctg	480
Db	64976	GTCACCCCTCCCTCGAGAGTATGCCCAAGTGTACAAACCGGTTCAAGGGCTGCATATGACCTG	6491
OY	481	atlgagatgaagcaagaaacagggattcccatcccggtcatctcccaagaaccggtgagc	540
Db	64916	ATGAAATAGGCAAGCAACATGGGGTTCCCATTCAGTCACTCCCAAGACCCGCTGAAG	6485
OY	541	atlgatltagaacctcatctgacatcagctcagaaggtctgtacatctcctgagaaccaccaagac	600
Db	64856	ATAGACAGAACTCATGTGCATTCAGCTATAGGCTGGAATCTCG--GAACCCAGAGAAC	6479
OY	601	caagcgcctccaggtctctcaacgaagaccgaaccgaacccccaaccct	660
Db	64798	CAAGTGCCTCCAGCTCTCATATGAAGACCCAGGACCCGAAGGCCCTCAACACCCCT	6473
OY	661	gacatctctgagatgagttcaaaaaaggggtccctctgtgaaggtgtgaccaacgtccaaggt	720
Db	64738	GACATTTCTGAGATCAAGTTCAAAAAGGGGTCCCGTGAAGGTGACCAACGCTCAAGAT	6467
OY	721	ggacaaccaccaacagacccctcctcttggagctctctcatgtactcgtgaacggaagtcgsggacaag	780
Db	64678	GGCACCAACCCACAGACCTCTCTTGAGCGTTTCGHTACTGATGAAGTTGCAAGGCAAG	6461
OY	781	catggtgtggccgtatgtgactctgtgagaacccgtctcatctgtgaatgtgaagttccgaagt	840
Db	64618	CATGGGCTGGGCGTATTTACATCTGTGAAGAACCGCTTCATTTGAATGAAGTCTTAGGT	6455
OY	841	atctacagagaccccaagagacacatcttaccatgtcatcttlaagatcatcgaggtctc	900
Db	64558	ATCTGAGAGGCCCAAGACAGAGGACATCTCTTACCACACACTATTTAAGCATCAAGGCTTC	6449
OY	901	accatgtgaccggyaagttgcgcaaaatcaacaagagccttggcgttgaaatttctgtgactg	960
Db	64498	ACCATGTACCGGAATGCGCAAAATCAACAAAGCCGTGAGCTTGAAATTTGCTGTGAGCTG	6443
OY	961	gttataccggttttaccgtctagcccttgaagtgtttaatcttccgcacactcgtcagcagaag	1020
Db	64438	GTTGTATACCGGTTTTCGSGATAGCCCTTAGTGTGAATTTGTCCGCCACGTATATGCCAAG	6437
OY	1021	tcccaagagcagttgtgaaaggtgcaagtgtctcaggttccctcaagggccaggtgtacatc	1080
Db	64378	TCCCAAGAAATGTAGTGAAGAGGAATGTGACGAGTGTCTCCCTCAAGGCCACGAGTATATC	6431
OY	1081	cttgcgcgggaattcccaactgtctctcaacatgagaagcttgttggacatgaacgtggcag	1140
Db	64318	CTTGCGTGGAGTCCCATCTGTCTCTTCAACAAGAGAGCTGGTGAACATGAACGTGGCAG	6425





60							*
40						*	*
20						*	*
0		*	*	*	*	*	*
	5	10	15	20	25	30	35
	Phrap Value Range						
	-----						

Version: 1.01 qxf0.

FEATURES	Location/Qualifiers
source	1. .159687

repeat_region	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="12" /clone="RP11-267D19" 11..542 /rpt_family="L1MEC" 34..137
STS	/standard_name="139279" 593..700 /rpt_family="L1MEC" 725..1027 /rpt_family="AluXs" 1056..1401
repeat_region	/rpt_family="L1MEC" 1419..1974 /rpt_family="L1ME1" 1985..2067 /rpt_family="MLT1-INTERNAL 2141..2246 /rpt_family="MLT1-INTERNAL"
repeat_region	/rpt_family="MLT1D" 2421..2938 3967..3992 /rpt_family="AT_rich" 4687..4824
repeat_region	/rpt_family="MIR" 5032..5098 /rpt_family="AT_rich" 5927..6111
repeat_region	/rpt_family="L1Pba" 6107..6578 /rpt_family="L1Pba" 6601..6953
repeat_region	/rpt_family="L1Pba" 6954..8357 /rpt_family="PTRS" 8358..8778
repeat_region	/rpt_family="L1Pba" 8779..9084 /rpt_family="AluBg" 9085..9710
repeat_region	/rpt_family="L1Pba" 9881..9929 /rpt_family="GA-rich" complement(11501..11569)
repeat_region	/rpt_family="MIR" complement(11571..12240) /rpt_family="MER21B" complement(12242..12407)
repeat_region	/rpt_family="MIR" complement(12510..12804) /rpt_family="AluXs"

Query Match	85.2%;	Score 1055.4;	DB 9;	Length 159687;
Best Local Similarity	92.4%;	Pred. No. 2.2e-209;		
Matches 1145; Conservative	0;	Mismatches 86;	Indels 8;	Gaps 3

Qy 1 atgtccgcaaaagctccgtgtgtcttgcctacagttgcgccttgcacccctgtgcatc 60  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 55610 ATGTCCAGAAAAGGCTCTGTGTTCTGGGCTACAGTGTGGGCGCTGGACACCTCCGCAATC 5555T

QY	61	ctctgctgagctgaagaaacagagctatgcagctcatctgctatcttcggccaacattgcccga	120
Db	55550	CTCTGCTAGCTGAAGAACAGACGCTTTGATGTCTATTCCTGCTACTCGGCCACATTAGCCAG	55491
QY	121	aagaaagacttcagagaagccagaaagagagactgaagctctggcgccaaagaagtgttc	180
Db	55490	AAGAAAGATTTCAGAGAAACAGAGAAAGAA - GCATTGAAGCTTTGGGGCCAAAGAGTTTC	55432
QY	241	gcactgataagaaacgcctacactctctggagcaactctcttcgaagccctgcacgcgcgc	300
Db	55371	GCATGTGATGAGGACCGGCTACCTCCCGGGCATTCTCTCACAGGCCCTCATCTGCCTCG	55312
QY	301	aaacaagtgaatctgcgccagcgggaggggcccagaatgtgtgccagggccccaagaa	360
Db	55311	AAACAAGTGAATAATTCGCCACCGGGGGGGCCAAAGTATGTGTCCCATGTGCGTACGGGA	55253
QY	361	aaggaggaagataagctccggttttagctcagctctctactactcgcggcccccagaataag	420
Db	55251	AAGGGGAGCATATAGTTCAGTTTGAGTCAACTGCTACTCTGCTGCGCTCCCGAGATTAAG	55192
QY	421	gtcatgtctccctcggagga tgccttgaattctacaacccggttccaagggcgcgaatgcctg	480
Db	55191	GTCATTGCTCCCTGGAGCATGCGCTGAGTTTACAAACCGGTTCAAGGGCCCAAGCAACCTG	55133
QY	481	atggagtagcgaagaagcaaacggygattcccatcccggtcactcccaagaacccgtgagc	540
Db	55131	ATGGAAATGCCAAAGACACACGGAGATTCCCATCTCGTGCTATCCCAAGAACCTGTGGAGAC	55072
QY	541	atgataagaaacctatgcacatacagcaagagccttgaaatccctggagaaccccaagaac	600
Db	55071	ATGAAACAAGAACTCTATGACATCAGCTACAGAGCTGGAATCCTCGAGAAATCCCAAGAAC	55012
QY	601	caagcgctcccaagctctctacagaagaaccagaagcccaagccaaagccccaaccct	660
Db	55011	CAACCGGCTTCAGAGTCTTACCCGAGAAACCCAGAGACC - - - - - GGCCCCAACACCCCT	54958
QY	661	gaacttcctgagaatgagttcaaaaaaggggtccctctggagagtgagcaacgctcaagagt	720
Db	54957	GAACTTCTGAGATTAGTT - - - - - AAAAAGGGGCTCCCGTGAAGGTGACTGCATCAAGGAT	54899
QY	721	ggacacacccacagacccctcctcttgagctctctatctatcactgaaagagctcggycgaag	780
Db	54898	GGCTCCACCTACACAGACCTCTTGGAGCTTTCATGTACTGAAACGAAGTCCGGGGCAAA	54839
QY	781	catggcgtggagccgataatgacatcgtggagaacgcgttaacttgaatgaagtccgaagt	840
Db	54838	CACGGCTGGGCGTACTACATCGTGAAGAACCGTTTATTTGAAGAAAGTGTCCAGGT	54779
QY	841	atctacagaaaccacagcagagcaactccttacaatgctcatcttaagacatcagagccctc	900
Db	54778	ATCTAAGAGGCCCCGGCAGAGCACTATCTTAAACAGCGTATTTAAGCATCAAGGCTTC	54719
QY	901	accatgagcccggaagtgcgcgcaaaatcaacaagaagccttgggcttgaaatttctgaagctg	960
Db	54718	ACCATGAGACTGGGAATGCACAAAATCAAAACAGCGCTGGGCTTGAATTTGCTGAGATG	54659
QY	961	gttataccggtttacgctcgtacagcccttgaagtgttaatttctccgcacatgcacgcgaag	1020
Db	54658	GTTATATCCGGTTTTCGGACACGCACTAGTGTCAATTTTGCCTGCGCACTCTTCCGCAAG	54599
QY	1021	tcccaagagcgaagtgaaggaagaatgtcagtgctcgttcctcaagagccaggtgtatcctc	1080
Db	54598	TCCCAAGAGCCAGGTGAAGGGAATGTCAGGTGCTCTTCAAGGGCCAGAGGTGTACATC	54539
QY	1081	ctcggccgggaatccccaatgctctctacaataagaagcttggttgagctgaacgtgagc	1140
Db	54538	CTCGCCGGGATGCCCATGTGCTCTTACATATAAAGCTGGTGAAGCGTGAACATTCAGCG	54479



OY 1141 ggtatgtagccacacgagcagcgggtatcaatcaatccatccatgagcctgaag 1200  
 |||||||  
 Db 54478 GGTGATTATGAGCCAAATGACTGCTTTCATCAACATCACTCCCTCAGGCTGAG 54419  
 |||||||  
 OY 1201 gaattcatcgtcccaagcaagtcactgccaatag 1239  
 |||||||  
 Db 54418 GAATATCATCATCTCCAGAGCAAGTCTCTCCCAATATG 54380  
 |||||||  
 RESULT 12  
 AC026357/c  
 LOCUS Homo sapiens chromosome 12 clone RP11-85503, WORKING DRAFT  
 DEFINITION  
 AC026357  
 AC026357.22 GI:12044426  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 204371)  
 Musny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,J., Bimaga,K., Blankenburg,K., Bonin,D., Bouck,J.,  
 Bowie,S., Brieleva,M., Brown,E., Brown,M., Bryant,N.P., Buhey,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C.,  
 Hollins,B., Homs,J.F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,C., Liu,W.,  
 Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
 Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
 Oguh,M., Okunoye,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuboan,I., Rolfe,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,  
 Sisson,I., Sodergren,E., Soneike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,P., Tellier,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasequez,L., Vera,V., Villalón,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Wellington,S., Williams,G., Williams,A., Wleceyk,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
 and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 204371)  
 Morley,K.C.  
 Direct Submission  
 Submitted (22-MAR-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jan 8, 2001 this sequence version replaced gi:12043803.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HASE  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; M77789  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-terminator Big Dye; 94% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 201671 bases at least Q40  
 Consensus quality: 204577 bases at least Q30  
 Consensus quality: 205669 bases at least Q20  
 Estimated insert size: 201838; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 9.7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 8 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 52254: contig of 52254 bp in length  
 52255 52354: gap of unknown length  
 52355 101848: contig of 49494 bp in length  
 101849 101948: gap of unknown length  
 101949 134385: contig of 32437 bp in length  
 134386 134485: gap of unknown length  
 134486 164948: contig of 30463 bp in length  
 164949 165048: gap of unknown length  
 165049 191727: contig of 26679 bp in length  
 191728 191827: gap of unknown length  
 191828 201957: contig of 10130 bp in length  
 201958 202057: gap of unknown length  
 202058 203210: contig of 1153 bp in length  
 203211 203311: gap of unknown length  
 203311 204371: contig of 1061 bp in length.  
 FEATURES  
 source  
 1. 204371  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-85503"  
 BASE COUNT 64581 a 39919 c 38928 g 60229 t 714 others

ORIGIN  
 Query Match 85.2%; Score 1055.4; DB 2; Length 204371;  
 Best Local Similarity 92.4%; Pred. No. 2.1e-209;  
 Matches 1145; Conservative 0; Mismatches 86; Indels 8; Gaps 3;

OY 1 atctcgaagaagggccggtgttctggtcctacagtggtgagcctggaacccctgcatc 60  
 |||||||  
 Db 158529 ATGTCCAGAAAGGCTCTGTGTTCTGGGCTACAGAGGGGCTGAGACCTCTGCAATC 158470  
 |||||||  
 OY 61 ctctgtgtgctgaaggaagaagcctatgacgtcattgctctactctgccaacattggtccag 120  
 |||||||  
 Db 158469 CTCCTGTACTGCTGAGGACAAAGGCTATATGATGTCATTGGCTACTGCGCAACTATGCGAG 158410  
 |||||||  
 OY 121 aaggaagactctgaggaagcagaagaagacgtgaagccttgaggtgggccaagaaggtgtc 180  
 |||||||  
 Db 158409 AAGGAAGATTTCACGAAACCGAAGAA-GCATTGAAGCTTGGGGCCCAAAAAGGTGTTTC 158351  
 |||||||  
 OY 181 attgagagatcgaagcagaagatttltgagagagattcattcctgacggcgcatcagtcagc 240  
 |||||||  
 Db 158350 ATTGAGAGATGTACAGCAGGAGATTGTGTGGAGAGATTCTGGCCAGCAGCATCAGTTTCAGC 158291  
 |||||||



```

Db 33680 GGTCAATGTCCTCCCTGGAGATGCCCAAGTCTTACAAACAGGTTCCAAAGGTCCCAATGACT 33739
Qy 480 gattgagtagcgaagaacgaacagggattcccaatcccggtactcccaagaacccgtgag 539
Db 33740 TATGGAACACAAAGCAACAGGGATTCCCATCCAGTCACTCCCAAGAACCTGTGAA 33799
Qy 540 catgatgagaacccatgagcatagctagaggtctggaatccctggagaaccccaaga 599
Db 33800 CATGGAACGAAACCTCATGTCAGATGAGAAATGAGGCTGGAATCTTGGAACCCCTAAAGAA 33859
Qy 600 ccaagcgctccaggctcctacacgaagaaccaggaaccaggaaccccaagaaccccaacaccc 659
Db 33860 CCAACGATTTTCCAGGTCTCTCCACGAACACCAGAGACCCGGCCAAAGCCCCCAAGACCC 33919
Qy 660 tgaactctcgagatcgagttcaaaaaaggggtccctgtgaagtgaacaaactcaagga 719
Db 33920 TGACATTCCTCGAGATCGAGTTCAAAATAGAGATCCCATGAAGAGACCAACGTCAGAGA 33979
Qy 720 tggacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 779
Db 33980 TGGACACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 34039
Qy 780 gcatggcgtgggagcgtatgagatcgtgagagaaccgctcacttgagaatgaagtcccgagg 839
Db 34040 GTACAGCTTGGGCGCTATTGACATCAAGAGACACCTTCACTGGAATGAGAGTCCCAAG 34099
Qy 840 tatcagaagaacacacacacacacacacacacacacacacacacacacacacacacacacacac 899
Db 34100 TATCATGATGACACCCAGACACACCATGTTACATGCTCATTTAGACATCAGAGGCTT 34159
Qy 900 caccatgagcgggaagatgagcgaatcaaacagaagcgtggtctggaatttctgagct 959
Db 34160 CACCATGGAACCGGGAGATGATCAAAATCAAAATGAAGGCTGGGCTTGAATTTGCTGAGCT 34219
Qy 960 ggtgataacggttaacggtcgtgagctgagttgaatttctgcacacgtcagcgca 1019
Db 34220 GGTGATACCGGTTTCTGGCACAGCTTGAAGTTTATTTTCCACCACTGCTATTTGCCAA 34279
Qy 1020 gtcccaagagcgaatggaagaagtgcaagtgctgcctcactcaaggcccaagtgatcat 1079
Db 34280 GTCCAGAGAGTGAAGTGAAGGAAAGTGCAGGTGCGCTTTCAAGGGCTGAGTACAT 34339
Qy 1080 cctggcgcggaagtcacacacacacacacacacacacacacacacacacacacacacacacacacac 1139
Db 34340 CCTAGGCGGAGAGT-CCTCTGTTTATCTTACACAGAGAGTGTGAGCATGAACGTCGA 34398
Qy 1140 gggatattatagcgaactgtagccacgggttcatcaacatcaatccctcaagctgaa 1199
Db 34399 GGTGATATATAGCCCAATCGATGCCACAGTTTATCATCAACTCAATTCCTCAGGCTGAA 34458
Qy 1200 ggaatataatcgtcctcagacgaagatcatcgtccaatag 1239
Db 34459 GGAATATATCATCTCCAGACGAAGTCACTGCAAAATAG 34498

```

```

RESULT 14
LOCUS      HUMAS3PS      1868 bp      DNA      PRI      31-OCT-1994
DEFINITION Human argininosuccinate synthetase pseudogene 3.
ACCESSION K01846
VERSION    K01846.1 GI:179058
KEYWORDS   argininosuccinate synthetase; processed gene; pseudogene;
SOURCE     Human genomic DNA, Charron-4A library of Lawn et al, clone psi-AS-3.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1868)
AUTHORS   Freytag,S.O., Bock,H.G., Beaudet,A.L. and O'Brien,W.E.
TITLE     Molecular structures of human argininosuccinate synthetase
          pseudogenes. Evolutionary and mechanistic implications
JOURNAL   J. Biol. Chem. 259 (5), 3160-3166 (1984)

```

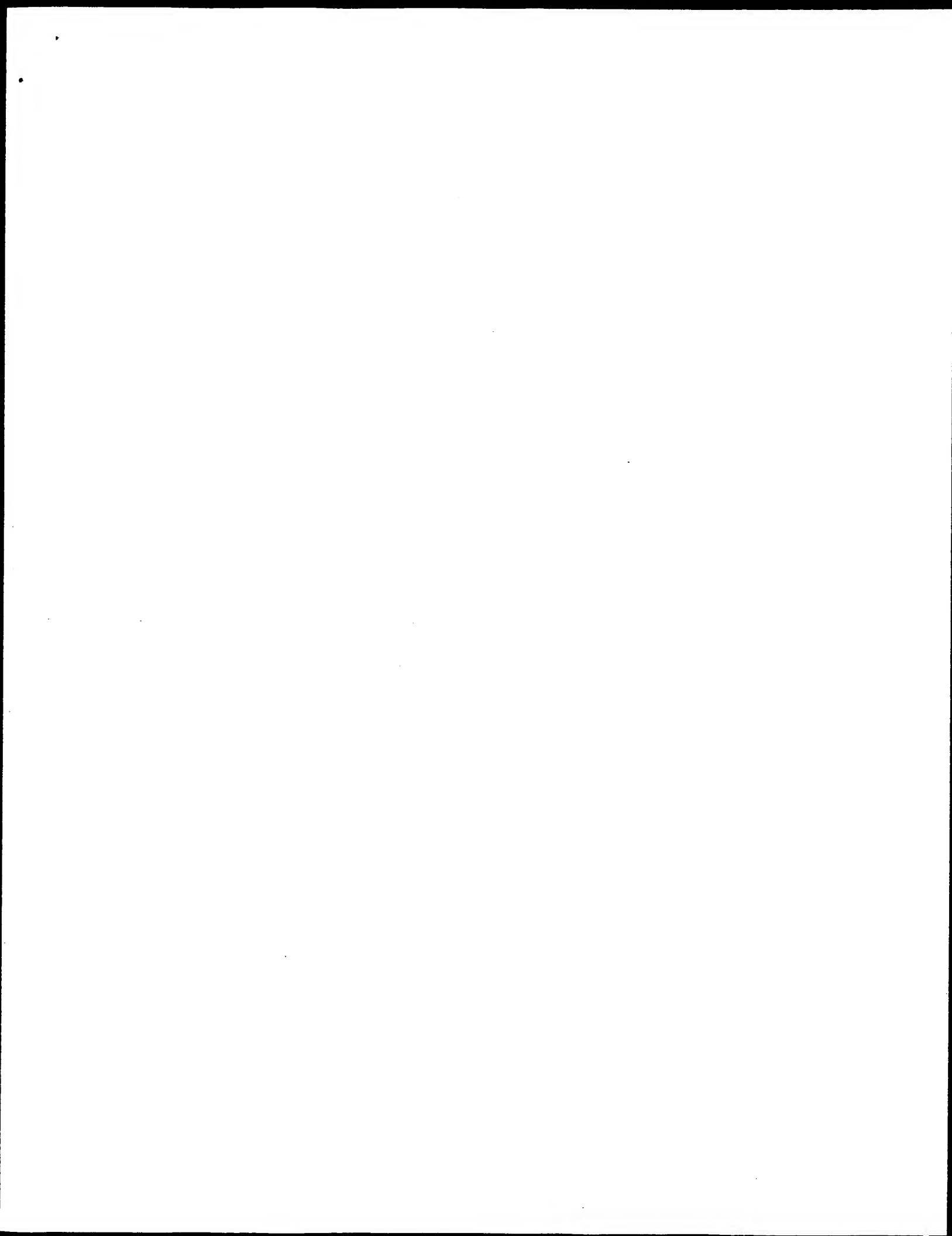
```

MEDLINE
COMMENT      84135824
              There is one expressed gene and 14 pseudogenes for argininosuccinate
              synthetase in the human genome. Three of the pseudogenes (psi-AS-1,
              -3, -7) were sequenced, in addition to the cDNA (AS). The
              pseudogenes are 89-93% homologous to the cDNA and lack introns. The
              authors [1] estimate that psi-AS-1 and psi-AS-3 were created 10-11
              million years ago (93% homology), and psi-AS-7 was created
              approximately 21 million years ago (89% homology). The functional
              gene contains 13 exons spanning 65 kb on chromosome 9 (9q34-qter
              locus).
FEATURES
    source          1..1868
                    Location/Qualifiers
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="9q11-q22"
    gene            1..1868
                    /gene="ASSP3"
    misc_feature    1..1868
                    /gene="ASSP3"
                    /note="500-119-709"
BASE COUNT      464 a 501 c 474 g 429 t
ORIGIN          329 bp upstream of a PstI site.
Query Match      84.9%; Score 1052.2; DB 9; Length 1868;
Best Local Similarity 92.3%; Pred. No. 2.2e-208;
Matches 1143; Conservative 0; Mismatches 88; Indels 8; Gaps 3;
Qy 1 atgtccagaagaagctccgtggtctggtcctacagtggtgcttgacacacctctccatc 60
Db 262 ATGTCCAGAAAGGCTGTGGTTGTGGCTTACAGTGTGGCTTGAGACACCTCTCCATC 321
Qy 61 ctctgtggtcgaaggaacaagcctatgacgtcattgtccctatctggtcccaacttgcag 120
Db 322 CTCCTGAGCTGAAGGAGCAAGGCTATGATGTCATTGTCCACTGACCAACATATACCA 381
Qy 121 aaggaagactctgaggaagcgaagaaggaagcactgaagcttggggccaagaagtgttc 180
Db 382 AAGGAAGATTTCAAGCAAAACACAGAAAGAA-GCATGTAGAGCTTGGGCGCAAAAAAGGTTC 440
Qy 181 attgagatctcagcagggagttctgtagagagatcatctatctgcgcgcacatccagc 240
Db 441 ATTGAGATCTCAGCAGGAGGATTTGTGAGAGATCATCTGGCCAGCCATCCAGTTCAC 500
Qy 241 gcaactgatatagaccgtctacccctctggtggaacctctctgcaagcgcctgcatcgccgc 300
Db 501 GCACCGGATGAGATCGCTACCTCCTGGGCACTTCTCCACAGGCGCCGCAATCGCTGC 560
Qy 301 aaacaagtggaatctgcccacgcggagggggccaaaglatgtgtcccaagcgccacaga 360
Db 561 AAACAAGTGGAAATTTGCCACCGGGAGGGGGCCAAAGTATGTGCCATGCGCTCAGGGGA 620
Qy 361 aaggggaacatagctcgtgtttagctcagctgactacactggtcccccaagtaag 420
Db 621 AAGGGAGCATAGGTCAGGTTGAGCTCAACTGCTACTCGTGGCTCCCAAGATTAAG 680
Qy 421 gtcatgtccctcggagagatcctgaattctacaacacgggttcaaggccgcaatgacctg 480
Db 681 GTCATGTCTCCCTGGAGCATATGCTGTGATCTCAACACGGTTCAAGGGCCACATGACCTG 740
Qy 481 atggagtagcgaagaacgaacgggattcccatcccggtcactcccaagaacccgtgagc 540
Db 741 ATGGAATAGCCAAAGACACACAGGATTTCCATTCCTGCTACCCCAAGAACTGTGGAC 800
Qy 541 atgagtagaacctctcgaacatcgaactcgaagcgtggaatccttggagaacccaagaac 600
Db 801 ATGAACAAGACCTCATGACATCACTACAGAGCTGGAAATCTTGAGAAATCCCAAGAAC 860
Qy 601 caagcgctccaggtctctacacgaagaccgaagccacgaagcccaacacccct 660
Db 861 CAAGCGCTCCCAAGTCTTACGCGAAGACCCAGAGATCC-----GACCCTCAACACCCCT 914
Qy 661 gaattctcgagatcgagttcaaaaaagggtccctgtgtaaggtagcaacgtcgaagat 720

```



```
|||||
Db 26758 GCACGTATGAGGACCACTACCTCTG6GTAATTCTCTACCAAGCCCTCATGCCCCAC 26817
OY 301 aaacaagtgaatacgcacgaagggggaagatgltgccacgycgcacagga 360
|||
Db 26818 AAAAAGTGGAAACCGTCCAGCGGCTGGGGCCCAAGTGTGTGCCACAGCCCAAGA 26877
OY 361 aagggaacgatacagtcggtttagctcagctcactcactcagcccccagataag 420
|||||
Db 26878 AAGGGGAACGATCAGGTCCAGTTCAGTCTGATCTGCTACTGCTGCCCCCAGATGAAG 26937
OY 421 gtc-----atgtcccttgagagatgcctgaattctaacacggttcaagggccgcaatg 475
|||
Db 26938 GTCCCTGAGAGGCTCCCTGGAGATGCCCGAGTCTACAACTGGTTCAAGGGCCGAGTG 26997
OY 476 accgtatggaatgaacgaagaacagcgattccatcccgctcactcccaagaaccgt 535
|||||
Db 26998 ACCTGATGGAATATGCCAAGACACAGGATTCCTCATCTGTGTCACCTCCAGAACCTGT 27057
OY 536 ggaacatgataagaaacctcactacacacacacacacacacacacacacacacac 595
|||||
Db 27058 GGAGATGAGACAGACAGACCTCATGCAATCAGCTACAAGGCTGGAATCCTGGAGAACCCA 27117
OY 596 agaaccgaagcctcccaaggtctctacacgaagaccgaagcccaagcccccaaca 655
|||||
Db 27118 AGAACCAAGTCTCCAGCTCTACACAAAGATCCAGGACCCAGCCAAAGCCCCCAACA 27177
OY 656 cccctgaattctgaagatcgagttcaaaaaaagggtccctgtaagtgtaacacagca 715
|||||
Db 27178 CCCCTGACATTTCTCAGATTCATTCAAAAAAGGGGTCCCTGAGGGTGCACAGCTCA 27237
OY 716 aggatgacacacacacacacacacacacacacacacacacacacacacacacac 775
Db 27238 AGGATGGACACACACACACACCTCTGAGCTCTTGCATGTACCTGAAAGAGTTGAG 27297
OY 776 gcaagcatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 835
Db 27298 GCACGACGAGTGTGGG-CTTATTTGCCATCTGTGAGAACATCTTCACCG-ACAAAGGCC 27355
OY 836 gaggatctacgaagacacacacacacacacacacacacacacacacacacacacac 895
Db 27356 AAGTATCTACCAAGACCCAGAGGACCATCTTACCATGCTCATTTAGACATCAAG 27415
OY 896 ccttacacatgagccggaagtgcgaaatacaaaacgaagccttggttgaattgctg 955
Db 27416 TCTTACCAAGACCCAGAGGACCATCTTACCATGCTCATTTAGACATCAAG 27475
OY 956 agctggtgtatccggtttagcgtcctacccctgaagtggtgaattggtcgacactgcatcg 1015
Db 27476 AGCTGATATACCCAGATTTCTGACATAGCCCTGAGTGTGAATTTGTCCGCCCTGCAATCG 27535
OY 1016 ccaagtcacgaagcgaatggaaggaagtgcaagtgctcgtcccaaggccaggtgt 1075
Db 27536 CCAAGTCCAGAGCGCGGACAGAGGAAGGAAAGTCCAGTGCCTCAAGGGCCAGGTGT 27595
OY 1076 acatccctcggccgggaagtcaccactgtctctacacatgaggaagctggtagcatgaaag 1135
Db 27596 ACATCTTGGCTGGAGTCCCATGCTCTCTACAAATGAGAGAGCTGATGAGCGTGAAGC 27655
OY 1136 tgcagggatattatagccaactgtagcaccgggttcatcaacataatccctcaaggc 1195
Db 27656 TGCAGGGTATATGAGCGGATGATGACCGGGGTTCATCAACATCACTCCTCAGAC 27715
OY 1196 tgaaggaatatactgtctcagagcaaggtcactgccaatatg 1239
|||||
Db 27716 TGAAGGAATACATTTGTCTCCAGATGAAGTTCACTGCCAGATAG 27759
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:42:27 ; Search time 146.6 Seconds

(without alignments)  
7245.743 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239

Sequence: 1 atgtccagcaaaagctccgt.....gcaaggtcactgcacaatag 1239

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID52/gcgdata/geneseq/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1232.6	99.5	1725	AAH72909	Human cervical can
2	1232.6	99.5	2810	AAH34821	Human colon cancer
3	1232.6	99.5	2812	AAH16219	Human prostate can
4	466.8	37.7	969	AAI12589	Probe #2522 for ge
5	466.8	37.7	969	AAI33937	Probe #2623 used t
6	466.8	37.7	969	AAI02494	Probe #2485 used t
7	450.4	36.4	678	AAI21777	Probe #11710 for g
8	450.4	36.4	678	AAI47060	Probe #15746 used t
9	450.4	36.4	678	AAI07462	Probe #7453 used t
10	351.4	28.4	1048	AAI14080	Aspergillus oryzae
11	310.4	25.1	360	AAI28968	Colon tumour relat

12	278.6	22.5	440	21	AAH98292	Human ovarian carc
13	229.2	18.5	234	21	AAH70071	Human secreted pro
14	227.2	18.3	270	21	AAH30247	Colon tumour relat
15	222	17.9	222	22	AAI29415	Colon tumour relat
16	217	17.5	217	22	AAI29308	Arabidopsis thalia
17	204.6	16.5	1653	21	AAH34397	Human colon cancer
18	166	13.4	175	21	AAH80429	Human colon cancer
19	166	13.4	175	21	AAH80726	Human colon cancer
20	158	12.8	269	22	AAH71231	Human cervical can
21	156	12.6	250	22	AAH72308	Human cervical can
22	143.8	11.6	259	22	AAH69430	C glutamylum codin
23	134.6	10.9	1203	22	AAH66511	C glutamylum codin
24	134.6	10.9	1209	22	AAH66512	C glutamylum codin
25	134.6	10.9	1326	22	AAH71920	Corynebacterium gl
26	134.6	10.9	1326	22	AAH68528	C glutamylum codin
27	114	9.2	127	19	AAH1829	Human biallelic po
28	96.4	7.8	1100	22	AAH52904	S. epidermidis ope
29	96.4	7.8	3153	22	AAH54077	S. epidermidis gen
30	65.6	5.3	855	18	AAH74785	Staphylococcus aur
31	56.2	4.5	400	18	AAH75150	Staphylococcus aur
32	50	4.0	92934	21	AAH81473	N. meningitidis pa
33	50	4.0	172325	21	AAH21613	N. meningitidis (s
34	44.6	3.6	1000	22	AAH91396	Streptococcus pneu
35	42.6	3.4	987	19	AAH52348	Aspergillus oryzae
36	41.8	3.4	874	21	AAH12057	Streptococcus pneu
37	40.6	3.3	7542	19	AAH52191	Homo sapiens mamma
38	40.4	3.3	2277	19	AAH13836	Human telomerase p
39	40.4	3.3	2277	19	AAH05372	Human secreted pro
40	40.4	3.3	2371	22	AAH91874	Human PC04397 nucl
41	40.4	3.3	2684	21	AAH58117	Human cDNA sequenc
42	40.4	3.3	3052	22	AAH19326	Corynebacterium gl
43	39.4	3.2	588	22	AAH71221	C glutamylum codin
44	39.4	3.2	1479	22	AAH68245	Viral-encoded sema
45	39.4	3.2	4707	20	AAH55560	

## ALIGNMENTS

RESULT 1	
AAH72909	
ID	AAH72909 standard; cDNA; 1725 BP.
AC	AAH72909;
XX	
DT	19-SEP-2001 (first entry)
DE	Human cervical cancer marker nucleic acid 4183.
XX	
KW	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX	
XX	Homo sapiens.
XX	
PN	WO200142467-A2.
XX	
PD	14-JUN-2001.
XX	
PF	08-DEC-2000; 2000WO-US33312.
XX	
PR	08-DEC-1999; 99US-0169681.
PR	21-DEC-1999; 99US-0171350.
PR	14-MAR-2000; 2000US-0189315.
PR	12-MAY-2000; 2000US-0203791.
PR	09-JUN-2000; 2000US-0210600.
PR	21-JUL-2000; 2000US-0220114.
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
PI	Schlegel R, Deeds J, Berger A, Zhao X;
XX	
DR	WPI; 2001-375006/39.
XX	
PT	New isolated nucleic acid for diagnosing and treating cervical cancer







PS Claim 1; Page 1093-1094; 2338bp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytostatic,  
 CC cardiactive, immunomodulatory, muscular, vulnery, gastrointestinal,  
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.

SQ Sequence 2812 BP; 594 A; 765 C; 863 G; 586 T; 4 other;

Query Match 99.5%; Score 1232.6; DB 21; Length 2812;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagcaaaagctccgtgtgtctgtgctacagtgagcgcttgacactctgcatc 60  
 DB 1297 atgtccagcaaaagctccgtgtgtctgtgctacagtgagcgcttgacactctgcatc 1356  
 QY 61 ctctgtgtgtgtgaagaacaaagctatgactgtctctctgtgccaatgtggcag 120  
 DB 1357 ctctgtgtgtgtgaagaacaaagctatgactgtctctctgtgccaatgtggcag 1416  
 QY 121 aaggaagcttcgaagaagcaagaagaagcactgaagcttggggccaaaagtgttc 180  
 DB 1417 aaggaagcttcgaagaagcaagaagaagcactgaagcttggggccaaaagtgttc 1476  
 QY 181 attgagatgtcgaagaagattgtgtgagagattcatctgtgcccagatccagtc 240  
 DB 1477 attgagatgtcgaagaagattgtgtgagagattcatctgtgcccagatccagtc 1536  
 QY 241 gcactgtatgaggaagcttactctctgtgcaactctctgtgccaagcctgtatgcgcgc 300  
 DB 1537 gcactgtatgaggaagcttactctctgtgcaactctctgtgccaagcctgtatgcgcgc 1556  
 QY 301 aaacaagtggaatcgcgcagcggaaggggcaagatgtgttcccaagcgccacagga 360  
 DB 1597 aaacaagtggaatcgcgcagcggaaggggcaagatgtgttcccaagcgccacagga 1656  
 QY 361 aaggggaacgatacagctcgtgtgtgactcagctgtactactctgtgccccagataag 420  
 DB 1657 aaggggaacgatacagctcgtgtgtgactcagctgtactactctgtgccccagataag 1716  
 QY 421 gtcatgtctcctgtgaggaatgctgtaattctacaacgggttcaaggcggaatgactg 480  
 DB 1717 gtcatgtctcctgtgaggaatgctgtaattctacaacgggttcaaggcggaatgactg 1776  
 QY 481 atgagatcgaagaacgaacgagatcccaatcccgctcactcccaagaacccgtgagc 540  
 DB 1777 atgagatcgaagaacgaacgagatcccaatcccgctcactcccaagaacccgtgagc 1836  
 QY 541 atgagatggaacatcagacatcagctgagagctgtgaatcctgtggaaccccaagaac 600  
 DB 1837 atgagatggaacatcagacatcagctgagagctgtgaatcctgtggaaccccaagaac 1896  
 QY 601 caagagctcgaagctctctacagaagaagccaggaaccccaagaaccccaacccct 660  
 DB 1897 caagagctcgaagctctctacagaagaagccaggaaccccaagaaccccaacccct 1956  
 QY 661 gacatctcgaagatcgaatcaaaaaaggggtccctgtggaagtgacaacagtcaagat 720  
 DB 1957 gacatctcgaagatcgaatcaaaaaaggggtccctgtggaagtgacaacagtcaagat 2016  
 QY 721 ggcaccaccacagacactctctgtgagctcttcatgttacctgaagaagtcgcgggcaag 780

DB 2017 ggcaccaccacagacactctctgtgagctcttcatgttactctgaagaagtcgcgggcaag 2076  
 QY 781 catggtgtggtccgtatttgacatcgtgtgagaacccgttcatgtgaatgattcccgaggt 840  
 DB 2077 catggtgtggtccgtatttgacatcgtgtgagaacccgttcatgtgaatgattcccgaggt 2136  
 QY 841 atctacagagcccaagaagcaccatcttaccatgacatcattgaatgaatcagagccttc 900  
 DB 2137 atctacagagcccaagaagcaccatcttaccatgacatcattgaatgaatcagagccttc 2196  
 QY 901 accatgaccggggaagtgcgcaaatcaaacgaagcctgtggttgaattgtctgagctg 960  
 DB 2197 accatgaccggggaagtgcgcaaatcaaacgaagcctgtggttgaattgtctgagctg 2256  
 QY 961 gttataccggttttaagcgtctagcctctgtgtgtgaattgttcgcacatcagtcgcaag 1020  
 DB 2257 gttataccggtttctgtgacagcctctgtgtgtgaattgttcgcacatcagtcgcaag 2316  
 QY 1021 tcccaagagcgaatggaagaagagtcgaagtgctccgtctccaaggccaagtgtaacatc 1080  
 DB 2317 tcccaagagcgaatggaagaagagtcgaagtgctccgtctccaaggccaagtgtaacatc 2376  
 QY 1081 ctgcgcgggagtcgcccaatgtctcttacaatgagagctgtgtgagcatgagctgtcag 1140  
 DB 2377 ctgcgcgggagtcgcccaatgtctcttacaatgagagctgtgtgagcatgagctgtcag 2436  
 QY 1141 gttgattatgagcgaactgtatgacacgggttcatcaaatatctctcagcgtcgaag 1200  
 DB 2437 gttgattatgagcgaactgtatgacacgggttcatcaaatatctctcagcgtcgaag 2496  
 QY 1201 gaatacatcgtctccagaagcaggtcactgtccaaatag 1239  
 DB 2497 gaatacatcgtctccagaagcaggtcactgtccaaatag 2535

RESULT 4  
 AAF12589  
 ID AAF12589 standard; DNA; 969 BP.  
 XX  
 AC AAF12589;  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Probe #5222 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe: human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 XX  
 PN W0200157278-A2.  
 PN  
 PD 09-AUG-2001.  
 PD  
 XX  
 PF 30-JAN-2001; 2001WO-US00670.  
 PF  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR  
 PR 26-MAY-2000; 2000US-0207456.  
 PR  
 PR 30-JUN-2000; 2000US-0608408.  
 PR  
 PR 03-AUG-2000; 2000US-0632366.  
 PR  
 PR 21-SEP-2000; 2000US-0234687.  
 PR  
 PR 27-SEP-2000; 2000US-0236359.  
 PR  
 PR 04-OCT-2000; 2000GB-0024263.  
 PR  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI  
 XX  
 DR WPI; 2001-488901/53.  
 DR  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -

xx Claim 25; SEQ ID No 2522; 487bp; English.

xx The present invention relates to human single exon nucleic acid probes  
 CC (SNP). The present sequence is one such probe. The SNPs are derived  
 CC from human HeLa cells. The SNPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

xx Sequence 969 BP; 249 A; 255 C; 244 G; 221 T; 0 other:

Query Match 37.7%; Score 466.8; DB 22; Length 969;  
 Best Local Similarity 88.0%; Pred. No. 6.9e-114;  
 Matches 521; Conservative 0; Mismatches 67; Indels 4; Gaps 1;

QY 1 atgtccagcaaaagctccgtgtctgtgctcagctgagcgctgagcaacctgtgcatc 60  
 Db atgtccagcaaaagctcagctgctgtgtgctcagctgagcgctgagcaacctgtgcatc 60  
 QY 61 ctctgt 120  
 Db ctctgt 120  
 QY 441 ctgt 500  
 Db ctgt 500  
 QY 121 aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 180  
 Db aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 180  
 QY 501 aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 560  
 Db aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 560  
 QY 181 atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
 Db atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
 QY 561 atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 620  
 Db atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 620  
 QY 241 gcaatgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 300  
 Db gcaatgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 300  
 QY 621 gccctgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 680  
 Db gccctgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 680  
 QY 301 aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
 Db aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
 QY 661 aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 740  
 Db aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 740  
 QY 361 aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 420  
 Db aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 420  
 QY 741 aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 796  
 Db aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 796  
 QY 421 gtcattgtccctgt 480  
 Db gtcattgtccctgt 480  
 QY 797 gtcattgtccctgt 856  
 Db gtcattgtccctgt 856  
 QY 481 atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 540  
 Db atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 540  
 QY 857 atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 916  
 Db atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 916  
 QY 541 atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 592  
 Db atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 592  
 QY 917 atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 968  
 Db atgaggaatgtcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 968

RESULT 5  
 AAI33937 standard; DNA; 969 BP.

XX AAI33937;

DT 17-OCT-2001 (first entry)

XX Probe #2623 used to measure gene expression in human placenta sample.  
 XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.  
 OS Homo sapiens.  
 PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 2623; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

XX The present sequence is one such probe. The probes are useful for

XX producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.

XX Sequence 969 BP; 249 A; 255 C; 244 G; 221 T; 0 other;

Query Match 37.7%; Score 466.8; DB 22; Length 969;  
 Best Local Similarity 88.0%; Pred. No. 6.9e-114;  
 Matches 521; Conservative 0; Mismatches 67; Indels 4; Gaps 1;

QY 1 atgtccagcaaaagctccgtgtctgtgctcagctgagcgctgagcaacctgtgcatc 60  
 Db atgtccagcaaaagctcagctgctgtgtgctcagctgagcgctgagcaacctgtgcatc 60  
 QY 381 atgtccagcaaaagctcagctgctgtgtgctcagctgagcgctgagcaacctgtgcatc 440  
 Db atgtccagcaaaagctcagctgctgtgtgctcagctgagcgctgagcaacctgtgcatc 440  
 QY 61 ctctgt 120  
 Db ctctgt 120  
 QY 441 ctgt 500  
 Db ctgt 500  
 QY 121 aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 180  
 Db aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 180  
 QY 501 aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 560  
 Db aaggaagcttcgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 560  
 QY 181 atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
 Db atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240  
 QY 561 atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 620  
 Db atgaggaatgtcagcaaggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 620  
 QY 241 gcaatgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 300  
 Db gcaatgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 300  
 QY 621 gccctgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 680  
 Db gccctgtatgaggaagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 680  
 QY 301 aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
 Db aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
 QY 681 aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 740  
 Db aacaaagtggaatcagcaaggaaggaaggaaggaaggaaggaaggaaggaagga 740  
 QY 361 aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 420  
 Db aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 420  
 QY 741 aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 796  
 Db aaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 796  
 QY 421 gtcattgtccctgt 480  
 Db gtcattgtccctgt 480  
 QY 797 gtcattgtccctgt 856  
 Db gtcattgtccctgt 856

QY 1 atgtccagcaagctccgtgtctctgacctacagctggcgccctggaacacctgtgcattc 60

XX









CC The present invention describes colon tumour associated proteins (I) and  
CC the polynucleotides (II) that encode them. (I) have cytostatic activity,  
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
CC (II) may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate colon tumour associated protein (TCAP)  
CC expression, such as colonic cancer. For example, (I) and (II) may be  
CC used to treat disorders associated with decreased expression by  
CC rectifying mutations or deletions in a patient's genome that affect the  
CC activity of TCAPs by expressing inactive proteins or to supplement the  
CC patients own production of them. Additionally, (II) may be used to  
CC produce the TCAP proteins, by inserting the nucleic acids into a host  
CC cell culturing the cell to express the protein. (II) and its  
CC complementary sequences may also be used as DNA probes in diagnostic  
CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
CC quantitate the presence of similar nucleic acids in samples, and  
CC therefore which patients may be in need of restorative therapy. (I) may  
CC also be used as antigens in the production of antibodies against TCAPs  
CC and in assays to identify modulators of TCAP expression and activity.  
CC Anti-(I) antibodies and antagonists may also be used to down regulate  
CC TCAP expression and activity. The anti-(I) antibodies may also be used  
CC as diagnostic agents for detecting the presence of TCAPs in samples  
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA184640 to AA129512  
CC and AA24494 to AA24523 represent nucleotide and amino acid sequences  
CC given in the exemplification of the present invention.

50 Sequence 360 BP; 84 A; 96 C; 103 G; 73 T; 4 other;

Query Match	25.1%	Score 310.4	DB 22	Length 360
Best Local Similarity	95.5%	Pred. No. 1.2e-72		
Matches 317	Conservative	2	Mismatches 13	Indels 0
			Gaps	0

QY	758	actgtgaacgaagctctgcggcgcaagcatgcgctggtggccgtatattgacatctgtgtgaagccgct	871
Db	1	accgtgaacgaagctgcgctgcgcaagcatgctgctgtggccgtatttgacatctgtgtgaagccgct	60
QY	818	tcattgtgaactgaagctccggaggtatctacgagaccccaagcagcaccatccctttacatg	877
Db	61	tcattgtgaactgaagctccggaggtatctctcgcgagaccccaagcagcaccatccctttacatg	120
QY	878	cctcatttgaacatcgagagccttccaccatgagacggcgagagatgcgcgcaaatcaacaagagcc	937
Db	121	cctcatttgaacatcgagagccttccaccatgagacggcgagagatgcgcgcaaatcaacaagagcc	180
QY	938	tgggctctgaaaattgcgcgagctcgtgtgtataccggttttaacggctacgcctctgaagtgat	997
Db	181	tgggctctgaaaatttgcgtgagctcgtgtgtataccggtctctcgtgcaacagccctgaagtgat	240
QY	998	tgtctgcgcactgcacatcgcccaagctccccaagcgagctgtgaagggagaaagtgcgaagtgcg	1051
Db	241	tgtctgcgcactgcacatcgcccaagctccccaagcgagctgtgaagggagaaagtgcgaagtgcg	3000
QY	1058	tcctcaagagcgccaggtgtatacctctgcgcg	1089
Db	301	tcctcaagagcgccaggtgtatacctctgcgcg	332

RESULT 12

ID AAC98292 standard; cDNA; 440 BP

AAC98292; AC

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:302

KM Human colon cancer: colon cancer antigen; diagnosis; detection;  
KM identification; cytostatic; cardioactive; neuroprotective; vulnary  
KM immunomodulatory; muscular; gynaecological; gastrointestinal;  
KM nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
KM neural disorder; immune system disorder; muscular disorder;  
KM reproductive disorder; gastrointestinal disorder; renal disorder;

KM infectious disease; cardiovascular disorder; ss.

05 Homo sapiens

PN WO2000055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587534/55.

XX

PT antigens, useful for the treatment, prevention, and diagnosis of colon

XX

CC AAC69799 AAC98763 encode the human colon cancer associated proteins',  
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioactive, muscular,  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnerary, nephrotoxic, antiflfective and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins  
CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.

5Q Sequence 440 BP; 116 A; 126 C; 115 G; 78 T; 5 other;

Query Match	22.5%;	Score 278.6;	DB 21;	Length 440;
Best Local Similarity	96.6%;	Pred. No. 3.2e-64;		
Matches 284; Conservative	0;	Mismatches 10;	Indels 0;	Gaps 0;

[illegible]

## RESULT 13

ID AAA70071 standard; CDNA; 234 BP.



XX	AA70071;
AC	
DT	07-NOV-2000 (first entry)
XX	
DE	Human ovarian carcinoma antigen polynucleotide SEQ ID NO:382.
XX	
KW	Human; ovarian carcinoma; ovarian cancer; therapy; diagnosis;
KM	tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.
XX	
OS	Homo sapiens.
PN	WO200036107-A2.
PD	22-JUN-2000.
XX	
PF	17-DEC-1999; 99MO-US30270.
XX	
PR	17-DEC-1998; 98US-0215681.
PR	17-DEC-1998; 98US-0216003.
PR	23-JUN-1999; 99US-0338933.
PR	24-SEP-1999; 99US-0404879.
XX	
PA	(CORI-) CORIXA CORP.
PI	Mitcham JL, King GE, Algate PA, Frudakis TN;
DR	WPI, 2000-431589/37.
XX	
PT	Immunogenic portion of an ovarian carcinoma protein and the nucleic acid encoding it, useful for the diagnosis, prevention and treatment of cancer, preferably ovarian cancer -
XX	
PS	Claim 18; Page 194; 299pp; English.
CC	The present invention describes an isolated polypeptide comprising an immunogenic portion of an ovarian carcinoma protein (or its variants).
CC	Ovarian carcinoma proteins, and polynucleotides encoding them, have cytosolic activity and can be used in gene therapy and vaccines.
CC	Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines are useful for the prevention, diagnosis and treatment of cancer,
CC	preferably ovarian cancer. AA69691 to AA70077 and AB12552 to AB12557 represent human ovarian carcinoma polynucleotides and proteins used in the exemplification of the present invention.
CC	
XX	
SQ	Sequence 234 BP; 47 A; 63 C; 63 G; 61 T; 0 other;
Query Match	18.5%; Score 229.2; DB 21; Length 234;
Best Local Similarity	98.7%; Pred. No. 3e-51;
Matches 231; Conservative	0; Mismatches 3; Indels 0; Gaps 0.
OY	662 acattctcgagatcgaagtcaaaaagggtgccgtaggaagtgacccaagtcgaagatg 721
DB	234 ACATTCTGGAGATGAGTTCAAAAGGGGTCCCTGTGAAGTGACCACACTCAAGATG 175
OY	722 gcaaccaccacagaacctctctgtgagctcttcatgttaacctgaagtcgcgggcaagc 781
DB	174 GCACCAACCACAGACCTCCTTGAGCTTTATCATGTACTGAACGAAGTCGCGGCACAAC 115
OY	782 atgagctggcgccgttttacacatcgtaggaagacgcgttatgtgaatgaagtcgcgaagta 841
DB	114 ATGGGTGGGCCGTATTGACATCGTGGAGAACCCTCTTCTTTGGATTGAATGCCGAGTA 55
OY	842 tctacgaagaccacagcaggaacactccttatacattgaatgaatgaatgaagtcgag 895
DB	54 TCTACGAGAACCCACAGCACACATCTTTACACAGCTCATTTAGACATCGAGG 1
RESULT 14	
AAC30247	
ID AAC30247 standard; cDNA; 270 BP.	
XX	

[illegible]

```

XX  AA129415;
AC  12-OCT-2001 (first entry)
XX  Colon tumour related determined cDNA sequence for clone R0097.C02.
XX  DE
XX  Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX  Immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX  Homo sapiens.
XX  W0200149716-A2.
XX  12-JUL-2001.
XX  29-DEC-2000; 2000WO-US35596.
XX  30-DEC-1999; 99US-0476296.
XX  10-JAN-2000; 2000US-0480321.
XX  15-FEB-2000; 2000US-0504629.
XX  06-MAR-2000; 2000US-0519444.
XX  19-MAY-2000; 2000US-0575251.
XX  29-JUN-2000; 2000US-0609448.
XX  28-AUG-2000; 2000US-0649811.
XX  (CORI-) CORIXA CORP.
XX  Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX  King GE, Wang T, Jiang Y;
XX  WPI: 2001-441847/47.
XX  Colon tumor associated proteins and nucleic acids useful for the
XX  prevention, diagnosis and treatment of colonic cancer -
XX  Claim 2; Page 400; 472pp; English.
XX  The present invention describes colon tumour associated proteins (I) and
XX  the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX  (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX  (II) may be used in the prevention, diagnosis and treatment of diseases
XX  associated with inappropriate colon tumour associated protein (TCAP)
XX  expression, such as colonic cancer. For example, (I) and (II) may be
XX  used to treat disorders associated with decreased expression by
XX  rectifying mutations or deletions in a patient's genome that affect the
XX  activity of TCAPs by expressing inactive proteins or to supplement the
XX  patients own production of them. Additionally, (II) may be used to
XX  produce the TCAP proteins, by inserting the nucleic acids into a host
XX  cell culturing the cell to express the protein. (II) and its
XX  complementary sequences may also be used as DNA probes in diagnostic
XX  polymerase chain reaction (PCR) and hybridisation assays to detect and
XX  quantitate the presence of similar nucleic acids in samples, and
XX  therefore which patients may be in need of restorative therapy. (I) may
XX  also be used as antigens in the production of antibodies against TCAPs
XX  and in assays to identify modulators of TCAP expression and activity.
XX  Anti-(I) antibodies and antagonists may also be used to down regulate
XX  TCAP expression and activity. The anti-(I) antibodies may also be used
XX  as diagnostic agents for detecting the presence of TCAPs in samples
XX  (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
XX  and AA24494 to AA24523 represent nucleotide and amino acid sequences
XX  given in the exemplification of the present invention.
XX  Sequence 222 BP; 31 A; 49 C; 76 G; 66 T; 0 other;

```

Query Match 17.9%; Score 222; DB 22; Length 222;

Best Local Similarity 100.0%; Pred. No. 2.4e-49;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY  446 aattctaacacggttcaaggcgcaatgacctgagtgtgagtgaagcaagcaacagcggga 505
DB  222 AATTCTAACACGGTTTCAAGGCGCAATGACCTGATGTGAGTACCAAGCAACACGCGGA 163

```

```

QY  506 ttcccatcccggtcactctcccaagaacccgttgagcaatgagtgaaccctcaccatca 565
DB  162 TTCCCATCCCGGTCTACTCTCCCAAGAACCCGTTGAGCATGATGAACTCTATGCACATCA 103
QY  566 gctacagagctggaatccttgagaagcccaagaacagcctccaggctctacacga 625
DB  102 GCTACGAGGCTGGATCTCTTGAGAACCCCAAGAACCAAGCGCTCTCAGGTCTACACGA 43
QY  626 agaccagagaccagcccaagaagccccaacacacccctgacatc 667
DB  42 AGACCCAGAGACCCAGCCAAAGCCCCCAACACCCCTGACATTC 1

```

Search completed: February 12, 2002, 13:12:38  
Job time: 1811 sec









Db 816 GAAGCGCCCCGCCCCGC 833

```

? APPLICANT: Melanie K. Spriggs Michael R. Comeau,
? APPLICANT: Robert F. Dubose, Richard S. Johnson
? TITLE OF INVENTION: VIRAL ENCODED SENAPHORIN PROTEIN
? TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
? NUMBER OF SEQUENCES: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Janis C. Henry
? STREET: 51 University St.
? CITY: Seattle
? STATE: WA
? COUNTRY: US
? ZIP: 98101
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/181,706
? FILING DATE: October 28, 1998
?
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/958,598 (converted to a
? APPLICATION NUMBER: Provisional, see below)
? FILING DATE: October 28, 1997
?
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: --to be assigned-- (USSN 08/958,598
? APPLICATION NUMBER: conversion to Provisional application)
? FILING DATE: October 26, 1998
?
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Henry, Janis C
? REGISTRATION NUMBER: 34,347
? REFERENCE/DOCKET NUMBER: 2631-A
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206)470-4189
? TELEFAX: (206)235-0644
?
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4707 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: linear
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..4707
?
? US-09-181-706-1

```



```

Patent No. 6174689
APPLICANT: Spriggs, Melanie
TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
RECEPTOR DNA AND POLYPEPTIDES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C. Henry
STREET: 51 University St.
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/458,791
FILING DATE: 10-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,598
FILING DATE: 28-Oct-1997
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4707 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4707
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-458-791-1

Query Match          3.2%; Score 39.4; DB 4; Length 4707;
Best Local Similarity 51.4%; Pred. No. 0.091; 86; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 125 aagactcgaaagaaagccggaagaagagcactgaaagcttggggccaaagaagtgtcatgtg 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 AGGACACGAGAGGGGGCGACGCTGGCCACGACAGAGACTGGGCGCCCTCAAGCTGTGCGAG 682

QY 185 aggaatgtcaagcaggagatttggaggaggtatatttggccggccatccagtcacagcgac 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 683 GCGCGGGCGCTGCACTTCTGTGGAGCGCTTCTCTGTGGAGCGGCAACATCTACTTCCCT 742

QY 245 tgtatgaagacgcctacccctcctgggaacccctcttgcacagcgccctgacagcccgca 301
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 743 ACTACCCCTACCAACTATAGAGCGGGGCTGCCACCGGGCTGGCCACGATGCGCGCA 799

RESULT 9
US-09-459-066-1
; Sequence 1, Application US/09459066
; Patent No. 6187909
; GENERAL INFORMATION:
; APPLICANT: Spriggs, Melanie
; TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
; RECEPTOR DNA AND POLYPEPTIDES
;

```

```

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Janis C. Henry
STREET: 51 University St.
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS/Windows 95
SOFTWARE: Word for Windows 95, 7.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/459,066
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/958,598
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)470-4189
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4707 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4707
US-09-459-066-1

Query Match          3.2%   Score 39.4; DB 4; Length 4707;
Best Local Similarity 51.4%; Pred. No. 0.091;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0.

Qy      125 aagacttcgagaaggcagaagaaggaacgtgaagcttggggccaaagaagtgtatcg 184
        ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      623 AGGACAGCGAGGGGCCAGCTGCACGACGAGAGCTGGGGCCCTCAAGCTGTGGAAG 682

Qy      185 agaatgtcagaagaggttgttgagaagftcatctggccgcgatccagtcagcgcaac 244
        ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      683 GCCGGGGCACCTGGCACTTGTGTGAGCCCTTTCTCTTGGAACGGCAGCATCTACTTCCT 742

Qy      245 tgcataagagccgctacctctcttggcacctctcttgcagagccctgcagccgca 301
        ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      743 ACIAACCCTCACTATACGAGCGGCGCTGCCACGGGCTGGCCAGCATGGCGCGCA 799

RESULT 10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
```

NUMBER OF SEQ ID NOS: 2  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 2 4403765  
 : LENGTH: 4403765  
 : TYPE: DNA  
 : ORGANISM: Mycobacterium tuberculosis  
 : FEATURE:  
 : OTHER INFORMATION: CDC 1551  
 : OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 : OTHER INFORMATION: represent a, t, c or g  
 : US-09-103-8404-2

Query Match 3.1%; Score 37.8; DB 4; Length 4403765;  
 Best Local Similarity 48.0%; Pred. No. 10; Mismatches 117; Indels 0; Gaps 0;

Matches 108; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
 QY 80 aagctatgacgtcattcctatctgccaacattgccaagaagaactcagaggaag 139  
 DB 1419659 AAGACGAAGCGCTTCATGTCAGATGGATGGATGCTGCTCCTCCTCAGCAGG 1419600  
 QY 140 ccaggaagaagcactgaagttgggccaagaaagtgttattgagagtcagcaggg 199  
 DB 1419599 TGGCCACCCAGCAAGCCAGCTGCTGTGCAAGAGCTGGCCAGCGGAACCGGACCG 1419540  
 QY 200 agttgtgagagattcattctgcccagccatccagtcagcagcactatagagacgct 259  
 DB 1419539 AGATCGCCGAGAGAGGCTCTCAGCAACCACTGACCACTAAGCAGCGCCTACTTGG 1419480  
 QY 260 accctctgggcaactctctctcagcagccctgacgcgcgcgaac 304  
 DB 1419479 TCGTCGACGACCAAGCGCTACTGCTCCGCAATACGCCAGCCAGC 1419435

RESULT 11  
 : US-09-221-298-66  
 : Sequence 66; Application US/09221298  
 : Patent No. 6284241  
 : GENERAL INFORMATION:  
 : APPLICANT: Xu, Jiangchun  
 : TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS  
 : FILE REFERENCE: 210121.471  
 : CURRENT APPLICATION NUMBER: US/09/221,298  
 : NUMBER OF SEQ ID NOS: 112  
 : SOFTWARE: FastSeq for Windows Version 3.0  
 : SEQ ID NO 66  
 : LENGTH: 204  
 : TYPE: DNA  
 : ORGANISM: Human  
 : US-09-221-298-66

Query Match 3.0%; Score 37.2; DB 4; Length 204;  
 Best Local Similarity 59.4%; Pred. No. 0.075;  
 Matches 63; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 108 caacatggccagaagaagacttcagagaagcagcaagaagcactgaagcttgaggc 167  
 DB 26 ctataagtcacatcagtagagcagcagcagccggaaggaagcccttaagcagcgggc 85  
 QY 168 caaagaagtgatcattgagagatgcagcaggaaggaagttgtgaggaag 213  
 DB 86 ggaacagagatctctcgaagagccgcgctggagaggaaggaag 131

RESULT 12  
 : US-07-951-715A-2  
 : Sequence 2; Application US/07951715A  
 : Patent No. 5625136  
 : GENERAL INFORMATION:  
 : APPLICANT: Koziel, Michael G.

APPLICANT: Desai, Nalini M.  
 APPLICANT: Lewis, Kelly S.  
 APPLICANT: Kramer, Vance C.  
 APPLICANT: Warren, Gregory W.  
 APPLICANT: Evola, Stephen V.  
 APPLICANT: Crossland, Lyle D.  
 APPLICANT: Wright, Martha S.  
 APPLICANT: Merlin, Ellis J.  
 APPLICANT: Launis, Karen L.  
 APPLICANT: Rotheisen, Steven J.  
 APPLICANT: Bowman, Cindy G.  
 APPLICANT: Dawson, John L.  
 APPLICANT: Dunder, Erik M.  
 APPLICANT: Pace, Gary M.  
 APPLICANT: Suttie, Janet L.  
 TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30B  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/951,715A  
 FILING DATE: 25-SEP-1992  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/772,027  
 FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: SPULLI, W. Murray  
 REGISTRATION NUMBER: 32,943  
 REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8615  
 TELEFAX: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3468 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 HYPOTHETICAL: NO  
 FEATURE:  
 NAME/KEY: misc-feature  
 LOCATION: 1..3468  
 OTHER INFORMATION: /product = "Full-length pure maize  
 OTHER INFORMATION: optimized synthetic Bt"  
 OTHER INFORMATION: /note = "Disclosed in Figure 3 as syn17.mze"  
 : US-07-951-715A-2

Query Match 3.0%; Score 37.2; DB 1; Length 3468;  
 Best Local Similarity 49.0%; Pred. No. 0.35;  
 Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1000 gtccgacactgcatgcacaaagtcacagagcgaatggaaggaagtgagagtcgcgc 1059  
 DB 2485 GACGACGACGCGCCGCTGGGCAACCTGAGTCTCTGAGAGAGAACCCCTGTGGCGAG 2544  
 QY 1060 ctcaagggcagtgatcattcctgagcaggaagtcacacattctctcaaatgagga 1119  
 DB 2545 GCCCTGGCCGCGTGAAGCGCGCGAGAGAGAGTGGCGGCAACGCGGAGAGCTGGAG 2604



FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pace, Gary M.  
 REGISTRATION NUMBER: 40403  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8582  
 TELEFAX: (919)541-8589  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3468 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 FEATURE:  
 NAME/KEY: misc-feature  
 LOCATION: 1..3468  
 OTHER INFORMATION: /product= "Full-length pure maize"  
 OTHER INFORMATION: optimized synthetic Bt"  
 OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn11.mze"  
 US-08-459-448A-2

Query Match 3.0%; Score 37.2; DB 2; Length 3468;  
 Best Local Similarity 49.0%; Pred. No. 0.35; Mismatches 103; Indels 0; Gaps 0;

Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
 QY 1000 gtccgcaactgcatgcgaagtcccaagagcgagtggaagtgcaaggtgtccgctc 1059  
 DB 2485 GACGGCCACAGCCCGCTGGCAACTGTGAGAGAAACCCCTGTGGCCAG 2544  
 QY 1060 ctcaaggcgaggtgtacatctctgcccggagagtcaccatgtctcttacaatgaagag 1119  
 DB 2545 GCCCTGGCCCGCGTAAGCGCCGGAAGAGTGCGCCGACCAACCGCGAAGCTGTGAG 2604  
 QY 1120 ctggtgagcatgaacgtgcaagggtattatgagccaactgagtcaccgggttcatcaac 1179  
 DB 2605 TGGGAGACCAACATGTGTACAGAGGCCCAAGAGACGTGTGACGCCCTTGTGTGAC 2664  
 QY 1180 atcaattccctcaagctgaag 1201  
 DB 2665 AGCCAGTAGACGCCCTGCAGG 2686

RESULT 15  
 US-08-459-448A-4  
 Sequence 4, Application US/08459448A  
 Patent No. 5859336  
 GENERAL INFORMATION:  
 APPLICANT: Kozziel, Michael G.  
 APPLICANT: Desai, Nallini M.  
 APPLICANT: Lewis, Kelly S.  
 APPLICANT: Kramer, Vance C.  
 APPLICANT: Warren, Gregory W.  
 APPLICANT: Ewola, Stephen V.  
 APPLICANT: Crossland, Lyle D.  
 APPLICANT: Wright, Martha S.  
 APPLICANT: Meilin, Ellis J.  
 APPLICANT: Launis, Karen L.  
 APPLICANT: Rothstein, Steven J.  
 APPLICANT: Bowman, Cindy G.  
 APPLICANT: Dawson, John L.  
 APPLICANT: Dunder, Erik M.  
 APPLICANT: Pace, Gary M.  
 APPLICANT: Suttie, Janet L.  
 TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
 TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: No. 5859336art1s Corporation

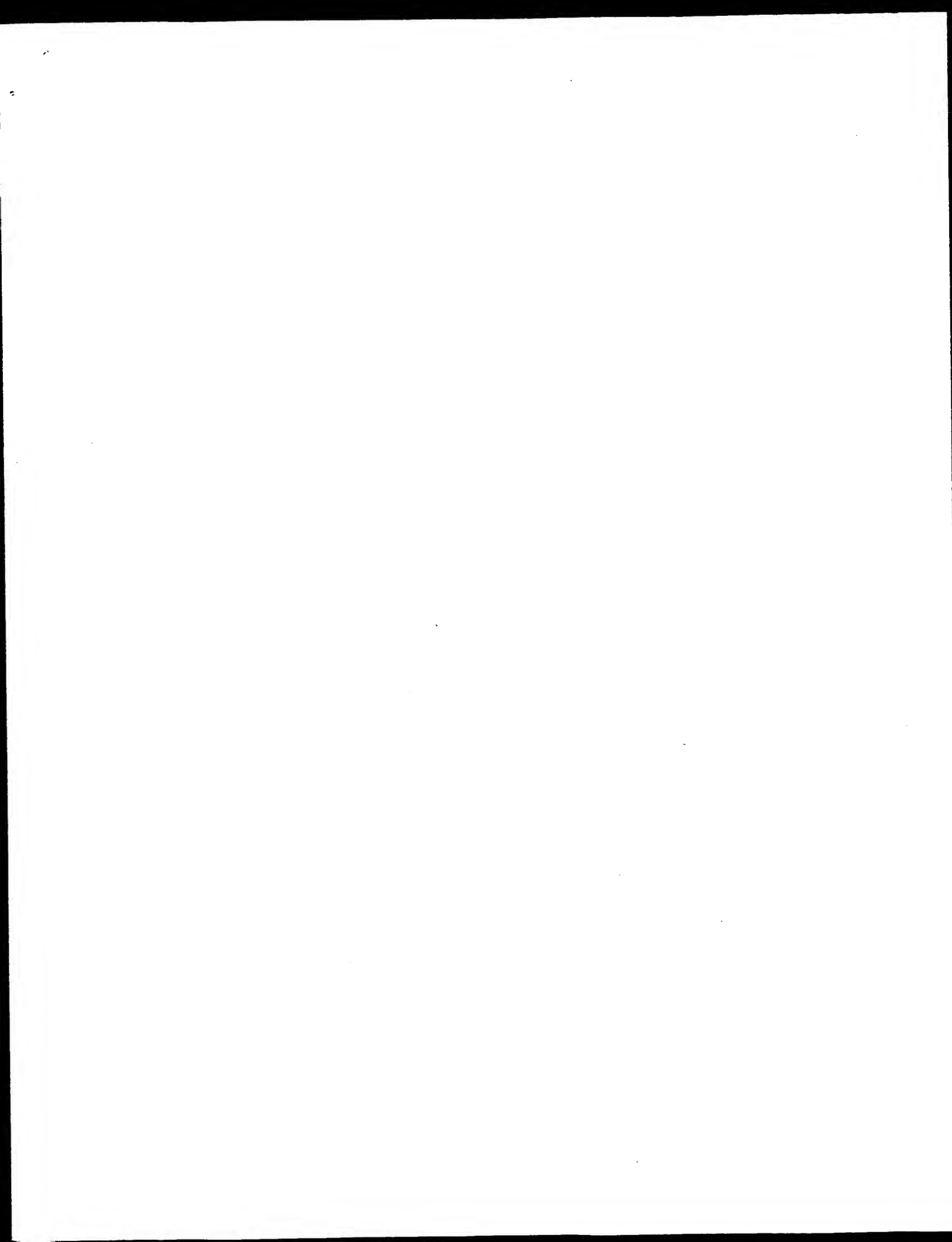
STREET: Patent & Trademark Dept., 520 White Plains  
 STREET: Rd., POB 2005  
 CITY: Tarrytown  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10591-9005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/459,448A  
 FILING DATE: 02-JUN-1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/951,715  
 FILING DATE: 25-SEP-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/772,027  
 FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pace, Gary M.  
 REGISTRATION NUMBER: 40403  
 REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIVA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3468 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "Synthetic DNA"  
 HYPOTHETICAL: NO  
 FEATURE:  
 NAME/KEY: misc-feature  
 LOCATION: 1..3468  
 OTHER INFORMATION: /product= "Full length synthetic"  
 OTHER INFORMATION: maize optimized"  
 OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn11.mod. This sequence  
 is identical to flsyn11.fin as disclosed in Figure 1."  
 US-08-459-448A-4

Query Match 3.0%; Score 37.2; DB 2; Length 3468;  
 Best Local Similarity 49.0%; Pred. No. 0.35; Mismatches 103; Indels 0; Gaps 0;

Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
 QY 1000 gtccgcaactgcatgcgaagtcccaagagcgagtggaagtgcaaggtgtccgctc 1059  
 DB 2485 GACGGCCACAGCCCGCTGGCAACTGTGAGAGAAACCCCTGTGGCCAG 2544  
 QY 1060 ctcaaggcgaggtgtacatctctgcccggagagtcaccatgtctcttacaatgaagag 1119  
 DB 2545 GCCCTGGCCCGCGTAAGCGCCGGAAGAGTGCGCCGACCAACCGCGAAGCTGTGAG 2604  
 QY 1120 ctggtgagcatgaacgtgcaagggtattatgagccaactgagtcaccgggttcatcaac 1179  
 DB 2605 TGGGAGACCAACATGTGTACAGAGGCCCAAGAGACGTGTGACGCCCTTGTGTGAC 2664  
 QY 1180 atcaattccctcaagctgaag 1201  
 DB 2665 AGCCAGTAGACGCCCTGCAGG 2686

Search completed: February 12, 2002, 14:41:32  
 Job time: 6940 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:41:27 ; Search time 1390.72 Seconds  
(without alignments)  
9573.472 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239  
Sequence: 1 atgtccagcaagaagctcgt.....gcaaggtcacgtcccaatag 1239

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: qd\_estl:\*  
11: qd\_est2:\*  
12: qd\_htc:\*  
13: qd\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	817.8	66.0	878	11	BI256461 602974488
2	809.8	65.4	878	11	BI258834 602869646
3	803.2	64.8	838	11	BI256559 602976921
4	783	63.2	875	11	BI257913 602970844
5	780.6	63.0	815	11	BI260468 602969267
6	780.2	63.0	980	10	BE379161 601238137
7	772.4	62.3	905	11	BE379161 601238137
8	768.8	62.1	918	11	BI259920 602968861
9	758.4	61.2	952	10	BE378581 601236939
10	753.2	60.8	981	11	BI085601 602870143
11	749.2	60.5	1028	11	BE206265 601869276
12	748.6	60.4	898	11	BI253774 602976122

13	745.4	60.2	851	11	BI258628 602969566
14	741	59.8	894	11	BI258439 602872821
15	738.8	59.6	851	11	BI255104 602975950
16	734.4	59.3	901	11	BI254017 602975178
17	733.4	59.2	822	11	BI258153 602971104
18	727.6	58.7	904	11	BE315954 601895904
19	726.6	58.6	1100	10	BE783168 601470778
20	725.6	58.6	900	11	BI253130 602873215
21	719.4	58.1	844	11	BI334157 602897591
22	717.6	57.9	766	11	BE543638 601071364
23	717.4	57.9	916	11	BE899161 601681957
24	715.8	57.8	1012	11	BE437839 602489758
25	715.4	57.7	843	10	BE614762 601281336
26	715	57.7	890	11	BI254203 602974855
27	707.6	57.1	1039	11	BE403394 602819070
28	705.2	56.9	906	11	BI334814 602989311
29	698.6	56.2	836	10	BE745306 601574086
30	696.6	56.2	1094	10	BE745308 601574086
31	696.2	55.9	862	11	BE394443 602456967
32	692.8	55.2	885	10	BE379436 601237366
33	692.2	55.9	1023	11	BE797765 601586321
34	692	55.9	731	10	BE745841 601580056
35	692	55.9	884	10	BE742766 601574370
36	690.8	55.8	820	11	BE400766 602464167
37	687.4	55.5	879	11	BE489990 602518529
38	686	55.4	733	11	BE741651 602631305
39	684.4	55.2	814	10	BE741651 602631305
40	682.2	55.1	763	11	BE792319 601585528
41	680.2	54.9	780	10	BE543488 601070608
42	680	54.9	820	10	BE543057 601068986
43	679.4	54.8	1013	10	BE378897 601237474
44	675.2	54.5	827	10	BE413326 60143326
45	675	54.5	971	10	BE615027 601281264

#### ALIGNMENTS

RESULT 1  
BI256461 878 bp mRNA EST 17-JUL-2001  
LOCUS 602974488F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5113965 5',  
DEFINITION BI256461  
ACCESSION BI256461  
VERSION BI256461.1 GI:14810890  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC <http://mgc.ncl.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgaabts-remail.nih.gov](mailto:cgaabts-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLM11276 row: 0 column: 22  
High quality sequence stop: 835.  
Location/Qualifiers  
1. 878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5113965"  
/clone\_lib="NIH\_MGC\_12"  
/tissue\_type="Cervical carcinoma cell line"  
/lab\_host="DH10B"

#### FEATURES

source





Db 361 AGGATCAGTCCGGTTGAGCTACGCTACTGAGCCGCCAGATAAGTCAATG 420

Qy 428 CTCCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 487

Db 421 CTCCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 480

Qy 488 ACCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 547

Db 481 ACCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 540

Qy 548 AATGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 607

Db 541 AATGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 600

Qy 608 CTCGAGTCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 667

Db 601 CTCGAGTCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 659

Qy 668 TCGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 726

Db 660 TCGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 719

Qy 727 ACCGAGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 786

Db 720 ACCGAGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 779

Qy 787 GTGGGCGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 846

Db 780 GTGGGCGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 837

Qy 847 GAGAGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 887

Db 838 AAGAGAGATGCTGAGAGATGCTGAACTTCTACAAACCGTGTCAAGGCGCCCAATGAGCTGATGAGT 877

RESULT 3

LOCUS B1256559 838 bp mRNA EST 17-JUL-2001

DEFINITION 602976921.F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5116133 5', mRNA sequence.

ACCESSION B1256559

VERSION B1256559.1 GI:14811085

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 838)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabp-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILMIL at:

http://image.llnl.gov

Plate: L1M11282 row: j column: 06

High quality sequence stop: 810.

location/Qualifiers

1. 838

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="NIH\_MGC\_12"

/lab\_host="cervical carcinoma cell line"

/lab\_type="cervical carcinoma cell line"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Library prepared by Life

BASE COUNT 212 a 237 c 239 g 150 t

ORIGIN

Query Match 64.8%; Score 803.2; DB 11; Length 838;

Best Local Similarity 99.4%; Pred. No. 1,5e-181;

Matches 827; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 3 gtcagcaagagcttcggtggttcgctacagtgagcgcctgagacacctcgtgacatc 62

Db 1 gtcagcaagagcttcggtggttcgctacagtgagcgcctgagacacctcgtgacatc 60

Qy 63 cgttgagctgaagagcaagagctatgacgctcattgctatcttgccaaacttgccagaa 122

Db 61 cgttgagctgaagagcaagagctatgacgctcattgctatcttgccaaacttgccagaa 120

Qy 123 ggaagacttcgagagagcagagagagagagagagagagagagagagagagagagag 182

Db 121 ggaagacttcgagagagcagagagagagagagagagagagagagagagagagagag 180

Qy 183 tgaagatgctgaag 242

Db 181 tgaagatgctgaag 240

Qy 243 actgatag 302

Db 241 actgatag 300

Qy 303 acaagtgag 362

Db 301 acaagtgag 360

Qy 363 gggag 422

Db 361 gggag 420

Qy 423 cattgctccctcgag 482

Db 421 cattgctccctcgag 480

Qy 483 ggaagtgag 542

Db 481 ggaagtgag 540

Qy 543 ggaagtgag 602

Db 541 ggaagtgag 600

Qy 603 aggccttcag 662

Db 601 aggccttcag 660

Qy 663 catttcgag 722

Db 661 catttcgag 720

Qy 723 caccacacacag 782

Db 721 caccacacacag 779

Qy 783 tggcgtgag 834

Db 780 tggcgtgag 830

RESULT 4

LOCUS B1257913 875 bp mRNA EST 17-JUL-2001

DEFINITION 602970844.F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5110307 5', mRNA sequence.

ACCESSION B1257913

VERSION B1257913.1 GI:14813752

Email: CGAP@bcrfemail.nih.gov, P.H.D.  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLM at:  
<http://image.hhl.gov>  
 Plate: LRAM1267 row: g column: 12  
 High quality sequence stop: 871.  
 Location: 20

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5110307"
/clone_lib="NH_MGC_12"
/tissue_type="Cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

```

Matches	837;	Conservative	0;	Mismatches	15;	Indels	5;	Gaps	4
---------	------	--------------	----	------------	-----	--------	----	------	---

6] ctcatatcaatcaaaacccccc  
20 atgtcctcagcamaaggctccgctggcttcttgacctacagtggcgccctggacacacccttcgtgatc 79

[illegible][illegible]

241 gcaatgtatatgaagacgcgtacctccctgycgacactcctcttaccagagcccttcattcgannccc 300  
|||||.....  
.....cccccccccccccccccccc

301 aaacaaagtggaatctgccacgcgaggggccaagatgtgtctccacgcgcacacaga 360

361 aagggaacatcagtcgcggtttgagctacgtcgtactactcaactgccccagataang 420

421 gtcattgctcccggaagatgtcctgaattctacaacggttcaaggccgcattgaactg 480  
|||||  
440 cttatcttct 480  
|||||

440 GTCATTCCTCCCTGGAGGATGCTGATTTCTACACCGGTTCAAGGCCGCATGACCTG 499

## RESULTS

<b>DEFINITION</b>	602969267F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108729 5'
<b>mRNA sequence</b>	UUG Up mRNA EST 17-JUL-2001

**KEYWORDS** EST.  
**SOURCE** human.

REFERENCE  
1 (bases 1 to 815)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NIN-MG bases 1 to 815

contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement, NINDS

Cloned and sequenced by: Incyte Genomics, Inc.  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information

```
http://image.llnl.gov
plate: LLAM11263 row: e column: 18
with snl144
```

```

1.013
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/tissue_type="cervical carcinoma cell line"
/lab_host="DH10R"

```

Average insert size 1.4 kb. Library prepared by Life Technologies.

204 a	236 c	232 g	143 t
-------	-------	-------	-------

Query Match	63.08;	Score 780.6;	DB 11;	Length 815;
Best Local Similarity	99.08;	Pred. No. 3.7e-176;		



```

|||||
Db 541 GGAGACGGCTTCATTGATGAACTCCCGAGGTATCTACGAGACCCCGACAGCCACCA 600
Qy 866 tcccttaacacatgcttaattagacatcgagccttcacatgagccggaagtgcacaa 925
Db 601 TCCCTTACCATTGCTATTGACATCGAGCCCTTCACCATGAGCCGGGAGAGTGGCCAAA 660
Qy 926 tcaaacagagcctgagcttgaattgctgaagctggtatatacaggtttacggcctaac 985
Db 661 TCAAAACAGAGCTGGGCTTGAATTTGCTGAGCTGTGTATACCGGTTTCTGGCAAGC 720
Qy 986 ctgaatgtaattgtccgcacatgcatc--gccaaagtcacagagcgaatg--aaagga 1042
Db 721 CTGAGTGAATTTGTCGCCCATCTGCATCCGCCAAGGTCCCAAGAGCAGTGGACGGA 780
Qy 1043 aagtcagagtgctgcctcctcaagggc---aagtgtaacatcctgcggcgaagctccac 1099
Db 781 AAGTCCAGGCTGTCCTCCCTCAAGGGGCGAGGTTCATTTCTCGCGCGGAGTCCCC 840
Qy 1100 tgcctctcaaatgagagagagagagagagagagagagagagagagagagagagag 1153
Db 841 CTGCTCTCTCTTACATGAGAGAGCTGTGTACCTTTCATGACGTGACGCTGATCTATGAG 900
Qy 1154 caactgagtcacacgaggtcatcaacatcaatcctcctcaagcctgaaggaat 1204
Db 901 CCACGTGATTCGCCCGGCTCATCAACACATCTTCCCGCAGTGAAGAGAT 951

```

```

RESULT 7
Bg828136 905 bp mRNA EST 22-MAY-2001
LOCUS 602753522F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4906132 5',
DEFINITION mRNA sequence.
ACCESSION Bg828136
VERSION Bg828136.1 GI:14175723
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 905)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHCN1808 row: 1 column: 05
High quality sequence stop: 790.
Location/Qualifiers
1..905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906132"
/clone_11b="NIH_MGC_17"
/tissue_type="Thadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pOTB7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT 227 a 265 c 253 g 160 t
ORIGIN
```

```

Query Match 62.3% Score 772.4; DB 11: Length 905;
Best Local Similarity 97.1% Pred. No. 3.5e-174;
Matches 851; Conservative 0; Mismatches 16; Indels 9; Gaps 6;

Qy 1 atcgcagcaaaagctcgtgcttcgcttaacagtcgagccttgagacactcgtgcatc 60
Db 27 ATGCCAGCAAAAGGCTCCGTTCTGGCTTACAGTGGCGGCTGGACACCTGCTCATC 86
Qy 61 ctggtgtgctgaagcaagcagcgtatgacgtatgctatcttgcccaatgtgcag 120
Db 87 CTCCTGTGCTGGAAGCAACAGGCTATGACGTCTATTCCTTCTGCAACATTTGGCCAG 146
Qy 121 aaggaagactctgaggaagcaggaagaaagcctgaagcttgaggccaa--aagtggt 179
Db 147 AAGGAAGACTTTCAGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 206
Qy 180 catgagagatgacagcagagagagagagagagagagagagagagagagagagagag 239
Db 207 CATTGAGAGATGTCAGCAGGAGGAGTTTGTGAGAGATTATCTGCGCCATCCAGTCCAG 266
Qy 240 cgaactgtatgagagacgcctactcctcgtgacactcctctgtcgaagcctgtca 299
Db 267 CGCACTGTATGAGAGACGCTACTCTCTGCGACCTCTCTTCCAGAGCCCTGCTATCCG 326
Qy 300 caaacagagtgaaatcgcccgagcagagagagagagagagagagagagagagagagag 359
Db 327 CAACAGAGTGAATTCGCCCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 386
Qy 360 aaaggaagacatcagatcagagagagagagagagagagagagagagagagagagagag 419
Db 387 AAAGGAGAGAGATGAGTCCGAGTTCGAGTTCAGTCTGCTGCTGCTGCTGCTGCTGCTG 446
Qy 420 ggtcattgctccctcgtgagagagagagagagagagagagagagagagagagagagag 479
Db 447 GGTATTGCTCCCTCGAGAGAGTGTGATTTCTCAACCGGTTCAAGGCGCGCATATACCT 506
Qy 480 gatgagtagcagaagcaagcaagcagagagagagagagagagagagagagagagagag 539
Db 507 GATGAGATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 566
Qy 540 catgagtagaagacatcagagagagagagagagagagagagagagagagagagagagag 599
Db 567 CATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626
Qy 600 ccaagcgtcctcgtgcttctacacagagagagagagagagagagagagagagagagag 659
Db 627 CCAAGCGCTCCAGGCTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 685
Qy 660 tgaatctcgaagtcagagagagagagagagagagagagagagagagagagagagagag 719
Db 686 TGACATTTCTGAGAGATGAGATTTCAAAACAAAGGGGTCGTGAGAGTGCACAGCTCAAGGA 745
Qy 720 tggcagcaccacacagacactccttgagagccttcacatgtaacgtgaagagtgaggcag 779
Db 746 TGGCAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 802
Qy 780 gcatgagcgttgagcgttatgacatcgtgagagagagagagagagagagagagagagag 839
Db 803 GCATGAGGCTGGGCGGTATTGACATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860
Qy 840 tatctagagagacccagcagagagagagagagagagagagagagagagagagagagag 875
Db 861 GATCTAGAGAAACCAAA--TTGGAGACATCTTTTACCA 894

RESULT 8
Bi259920 918 bp mRNA EST 17-JUL-2001
LOCUS 60296896F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108463 5',
DEFINITION mRNA sequence.
ACCESSION Bi259920
```

VERSION BI259920.1 GI:14817719

KEYWORDS

EST.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

SOURCE

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

ORGANISM

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

REFERENCE

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

AUTHORS

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

TITLE

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

JOURNAL

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

COMMENT

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

FEATURES

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

source

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

BASE COUNT

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

ORIGIN

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Query Match

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Best Local Similarity

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Human.

Matches 876; Conservative

Human.

Query Match	60.8%	Score 753.2	DB 11	Length 881
Best Local Similarity	96.2%	Pred. No. 1,3e-169		
Matches 825	Conservative	0	Mismatches 28	Indels 5
				Gaps
QY 1 atgcacagcaaaagctcgtgtgtcttgctctacagtgcgaccttgacacctctgtgcac				82
23 ATGTCACGACAAAGCTCCGGTGTCTGGCTTACAGTGGACGGCTGGACACCTGTGGTAC				82
QY 61 ctctgtgtgctgaaggaacaaagctatgaagtcatgtgctctatcttgcccaacttggcag				12
23 ATGTCACGACAAAGCTCCGGTGTCTGGCTTACAGTGGACGGCTGGACACCTGTGGTAC				14
QY 83 CTCGTGTGGTGAAGGAAACAAAGCTTTGACGTCATTTGGCTTTCTGTGGCAACATTTGGCAG				14
121 aaggaagacctcggaagacaggaagaaaggaactgaagcttgggggccaagaagtgttc				18
143 AAGGAAGACTTTCGAGAGAAAGCCAGAGAAAGCCACTAAACTTTGGGGCCAAAAAAGTTTC				20
QY 121 aaggaagacctcggaagacaggaagaaaggaactgaagcttgggggccaagaagtgttc				20
143 AAGGAAGACTTTCGAGAGAAAGCCAGAGAAAGCCACTAAACTTTGGGGCCAAAAAAGTTTC				20
QY 143 AAGGAAGACTTTCGAGAGAAAGCCAGAGAAAGCCACTAAACTTTGGGGCCAAAAAAGTTTC				20
DB 181 atgtgaagatgtcagcaggagattgttgagaaatcatcttgcgagccatccatccagc				20
203 ATTGACGATGTACGACGGAAGTTGTGTGAGAGATTCATCTGTGGCCGGCATTCAGTCAAGC				20
QY 203 ATTGACGATGTACGACGGAAGTTGTGTGAGAGATTCATCTGTGGCCGGCATTCAGTCAAGC				20
DB 241 gacacttatgagagacgcgtacctccctctggcaactctcttgcagagccctgcatactccgc				3
263 GCACCTGTATCAGGACCGCTACCTCTCTGGGCAKCTCTTTCGACAGGCCCTGCAATCCCCCGC				3
QY 263 GCACCTGTATCAGGACCGCTACCTCTCTGGGCAKCTCTTTCGACAGGCCCTGCAATCCCCCGC				3
DB 301 aaacaatgtgaatctgccacgcggaagaggggccaagatattgtgtccacagggccacaga				3
323 AAACAGTGAATTCGCCCGACGGGAGAGGGGCCAAGTATGTGTGCCACGGCGCACACAGA				3





Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution  
 found through the I.M.A.G.E. Consortium/BLND at:  
<http://image.llnl.gov>  
 Locus: ULM1260 row: 1 column: 06  
 High quality sequence spot: 815.  
 Location/Qualifiers

```
Location/Name="
1. .898
/Organism="Homo sapiens"
```

229 a 251 c 255 g

	Score	DB	Length	Gaps
Query Match	60.48;	748.6;	11;	988;
Past Local Similarity	96.7%;	Pred. No. 1.7e-168;		

	Matches	80%: Conservative	0%: MSimilarches	100%: MSimilarches
QY	1	atgttcagcaaaagctctcgtgtctctgagctcaagtgcgcgccttgcgaacctgtgtc	60	
Db	18	ATGTCACAGCAAAAGCTCTCGTGTCTTGCGCTCAAGCGGGGCTTGAGACCTCGTGATC	77	
QY	61	ctcgtgtgcgtgaagaacaaagctatgcacgtatctgtccaccttgcgaacatgtgcag	120	
Db	78	CTCGTGTGCTTGAAGAAACAAGCTTAGTACGTCACTTGCCATCTGGCCCAACATTGGCCAG	137	
QY	121	aaggaagaccttcgaaggaagccagaagaagcacttgaagctttggagcccaaaagttgtc	180	
Db	138	AAGGAAGACTTCGAGGAAGCAAGCAAGGCACTGAACTTGAGGCGCAAAAAGGCTTC	197	
QY	181	attgcagagatgcagcagagagttctgtgaagagttcaatctgcgcgcacatccagttccagc	240	
Db	198	ATTGAGGATATCCAGCGAGGAGG- TTGTGGAGGGAGTTCAATCTGGCGGSCCATCAAGTCCAC	256	
QY	241	gcactgtatgaagaccgctactactcctctgggagacctctcttcgcaagccctctgcacgcgcg	300	
Db	257	GCACGTATAGAGGACCGCTCACTCTCGGGCACTCTTGTGGCAAGGCTCTCATTCGCCCGG	316	
QY	301	aaaagaatggaatacgcgccagcggagggggccaaagtatgtgtccacagcgccacacaga	360	
Db	317	AAGCAAGTGGAAATCGCCACGCGAGGGGGGCCAAGTATGTGTCTCCACGGGCGCACAGGA	376	

RESULT	13	17-JUL-2001
B1258628	851 bp	EST
LOCUS	602869566F1 NIH_MGC_12	Hom sapiens CDNA clone IMAGE:5109064 5',
DEFINITION	mRNA sequence.	
ACCESSION	B1258628	
VERSION	B1258628.1	GI:14815170

VERTEBRATA: Vertebrata; Euteleostomi;  
KEYWORDS EST. human.  
SOURCE Homo sapiens  
ABSTRACT

ORGANISMS: Metazoa; Chordata; Craniata; Vertebrata; Eukaryota; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 851)

**AUTHORS** NIH-MGC <http://mgc.nhl.nih.gov/>  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**EDITOR** Robert Strausberg, Ph.D.

COMMENT

Contact: Robert Schaefer, 505-845-1100  
Email: cgsabbs-remail.nih.gov  
Tissue Procurement: ATCC  
and Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.  
CDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing by: Incyte Genomics information can be  
clone distribution: MGC clone distribution information at:

found through the I.M.A.G.E. Consortium/<http://image.llnl.gov>  
 Plate: LLAM1264 row: c column: 17  
 High quality sequence stop: 849.

FEATURES	Location/Qualifiers
source	1. .851
	/organism="Homo sapiens"
	"0606"

```

/db_xref="taxon:3006
/clone="IMAGE:5109064"
/clone_lib="NIH_MGC_12"
/clone_type="cervical carcinoma cell line"

```

```

/tissue_type= cervix
/lab_host="DH10B"
/Vector= pCMV-SPORT6; Site_1: Not
/note="Organ: cervix;
```



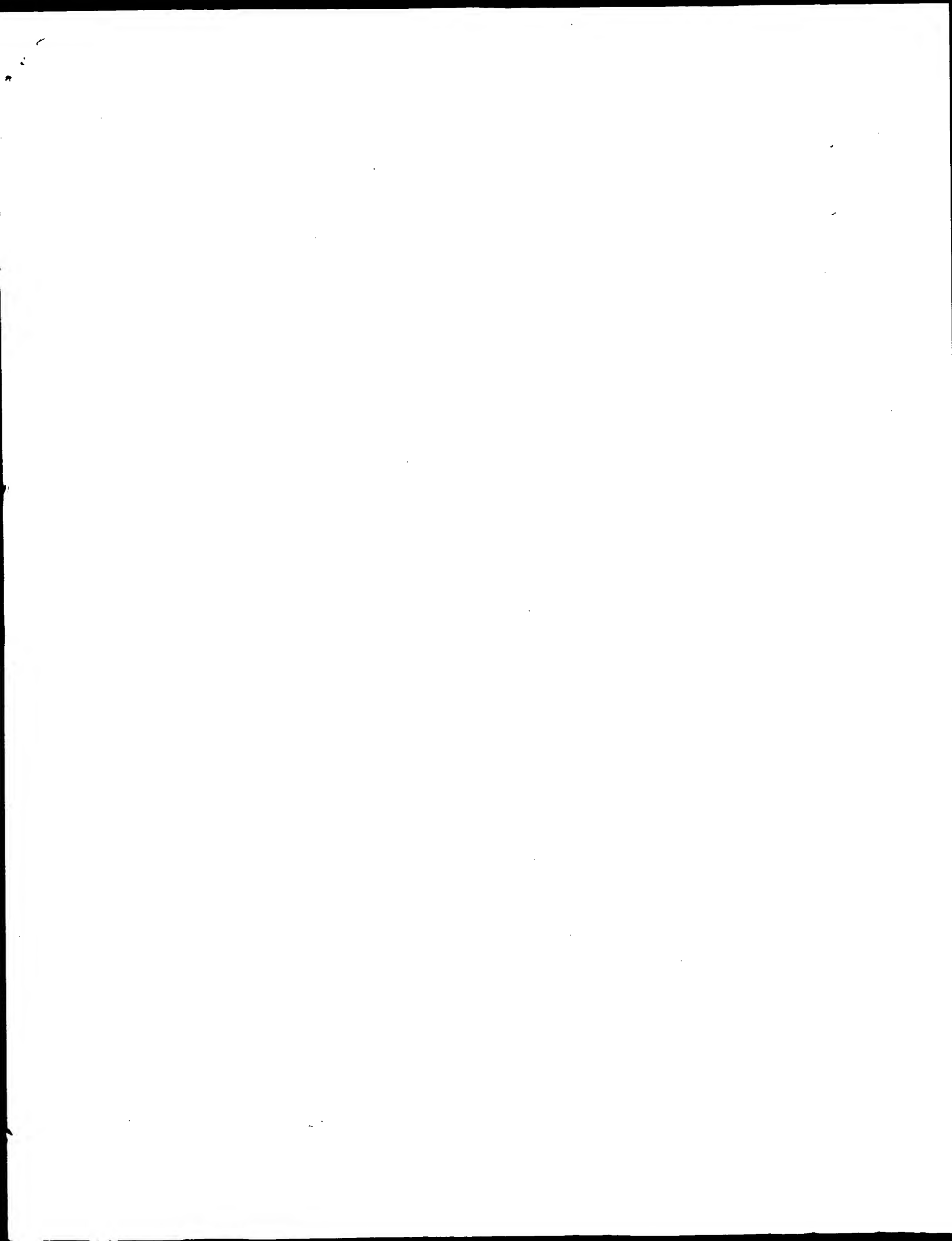
## RESULT 14

Query Match	59.8%	Score 741	DB 11	Length 894
Best Local Similarity	95.2%	Pred. No 1,1e-166		
Matches 829	Conservative 0	Mismatches 35	Indels 7	Gaps 6
Oy	1 atgtccagcaaaagctccgtgtgtctgtcctgaacgtgtgcgccttgaaacgttgatc	60		
Db	26 AtgtCCAGCAAAAGGCTCCGTGGTCTTGCCCTACAGGCGGCGCTGACACCTGGATC	85		
Oy	61 ctctgttggctgaagaagcaaggtactgaagtatgtcctatctctgtgccaactgtgcccag	120		
Db	86 CTCTGTGTGCTGAAGACAAGGCTATGACGTATGTGCTTCTGTGGCAACATTTGCCCGAG	145		
Oy	121 aaggaagacttcctggagaagcagaagaaggaacgtgaagctgtggtggccaaaaggtgttc	180		
Db	146 AAGCAAGACTTTCGAGAGACCAAGCAAGGAAGGACACTGAAGCTGTGGGCCAAAAAGGCGTTC	205		
Oy	181 attgaggaatgttcagcagaggaattgttgagaagttaactgtgcgcgcactcaagtccagc	240		
Db	206 ATTGAGGATGTCCAGCGGAGATTGTGGAGGAGTTCAATCTGGCGCGCATCCAGTCAAGC	265		
Oy	241 gcactgatatgaagccgcctactcctctgtgaacctctcttgcagagcccttgatctgcgcg	300		
Db	266 GCACGTATGAGAGACCGCTACCTCTCTTGCGCACCTCTTGGCAGGCCCTGTGATGCCCGC	325		
Oy	301 aaacaagtggaaatctgcaccagcgggaaggggccaagatagtgtcccaagggcgcacagga	360		
Db	326 AAMCAGTGTGAATGCGCCACCGGAGGGGCCAAGTATGTGTCCACAGCGCCACAGCA	385		
Oy	361 aaggggaacagatcaggtctggttttgagctcaactgtcacttaactcgtgcgccccagatanaag	420		
Db	386 AAGGGAACAGATCGAGTCCGGTTTGAGCTCAACCTGTACTACTAGCGCCCCAGATTAAG	445		

BASE COUNT	215 a	241 c	240 g	155 t
Query Match	59.6%;	Score 738.8;	DB 11;	Length 851;
Best Local Similarity	95.6%;	Pred. No. 3.6e-166;		
Matches 816; Conservative	0;	Mismatches 27;	Indels 11;	Gaps 5
8 gcaaaagctccggtgtctctgacaaagctgagcctgtgacacccctgtgacatccctgctgt	67			
1 GCAAAGGCTCCGTGTCTTGTGCTTACAGTGGCCGCTGACACCTGCGATCCTCGT	60			
68 gactgaagaacaaagctatgacgctcatctcctatctgacacatctgacgaagaagag	127			
1 GCAAAGGCTCCGTGTCTTGTGCTTACAGTGGCCGCTGACACCTGCGATCCTCGT	60			
61 GGTGAAGGAAACAAAGCTATGACGTCATTGGCTATCTGCGCCGCAATTCACAGGCGCACTGT	240			
128 actctgaggaagaacgaagaagaagactgaaacttgaggccaaagaagtgctaatgag	187			
121 ACTTCGAGGAAGCCACGAGAGAGCCACTGAAGCTTGGGGCCAAAMAAAGTGTTCATTGAGG	180			
188 atgtcaagcagagagttgttgaggaagttcatctatctgacgagccatccagtcagacagt	247			
181 ATGTCAAGAGGAACTTGTGAGAGATTCATCTGCGCCGCAATTCACAGGCGCACTGT	240			
248 atgagagcagctacccctcctggaactctctctgacagccctgacatcgccgcaaaag	307			
241 ATGAGAGCCGCTACCTCTGCGCCACTCTTCTTCCAGGCGCCGCTGACAGGCGCACTGT	240			
308 tggaaatcgcccaagcag	367			
301 TGGAAATCCGCCAGGGGGGAGGGGCGCAAGTATGTGTCCAGGCGCCACAGAGAGGGGA	360			
368 acgatcagctcgggtttgagctcagctgctacatctgaccccccagatgaagatcgt	427			
361 ACGATCAGGCTCCGTTGAGCTACACTCTCTTCCAGGCGCCGCTGACAGGCGCACTGT	420			
428 ctccctgaggaagctcgtgaatcttacaaccggttcaagagcgcagatgacctgtgaggt	487			
421 CTCCCTGAGAGATGCTCTGAATTCCTACACCGGTTCAAGGCGCGCAATGACCTGATGAGT	480			
488 acgcaagcaacacagagatcccatcccggtcactcccaagaacccgtgagcagtga	547			
481 ACGCAACCAACACAGGGATTCCTCCAGGCTCCAGAGACCCGAGGAGCATGTGATG	540			
548 agaacctcatgacatcagctcagagagctgaaatctctgagaaaccccaagaacacagcgc	607			
541 AGAACTCTCTGACATCACTACAGAGCTGAAATCTCTGGAGAA-CCCAAGAACCAAGCGC	599			
608 ctccagctctctacacgaagaacccaagcccaagcccaacacacccctgacatc	667			
600 CTCACAGTCTCTACAGAGAGCCCGAGCCCAAGCCCAAGGCCCCAGAA-CCCTGACATTC	658			
668 tcaagatcgaatccaacaaagagctcctgtgaaagtgaacaaagctgaagatgacacga	727			
659 TCGAGATCGAGTCAAAAAAGGGGT-CCGTGAGGCTGACCAACAGCTCAAGATGATGACCA	717			
728 ccaacacagacccctctgagagctcttcatgctacctaagaagctcgaggcaaga-ctgc	786			
718 CCACCAACACCTCTGAGCTCTCTCACTGACCTGACAGAGTCCGGGCAAGCATTTGGC	777			
787 gtggagcgtatgacatcgttggagaagcgttcaacatggaagctgaagctgagag	839			
778 GTGGGCGCTGATGACCATCGGTTGGCAGCAAGCTGTTCACCTTGGAATTGAAGTCCCAAG	837			
840 tatctcagagacc 853				
838 TATATACGAAAC 851				

Search completed: February 12, 2002, 13:10:00  
Job time: 1713 sec





Davis, M.  
09/775693  
Seq. ID 1 w/ Interf  
Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 12:46:27 ; Search time 2570.39 Seconds  
(without alignments)  
8448.217 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239  
I atgtccagcaagagctccgt.....gcaaggtcactgccaatag 1239

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 8763200856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending Patents\_NA\_Main.\*  
1: /cgn2\_6/ptodata/2/pna/US006.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US007.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US008.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US009.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US010.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US011.COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US012.COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US013.COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US014.COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US015.COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US016.COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US017.COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US018.COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US019.COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US020.COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US021.COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US022.COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US023.COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US024.COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US025.COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US026.COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US027.COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US028.COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US029.COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US030.COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US031.COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US032.COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US033.COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US034.COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US035.COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US036.COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US037.COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US038.COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US039.COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US040.COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US041.COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US042.COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US043.COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US044.COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US045.COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US046.COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US047.COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US048.COMB.seq.\*

Result

No.	Score	Query Match	Length	DB	ID	Description
1	1239	100.0	1239	30	US-09-775-693-1	Sequence 1, Appl
2	1232.6	99.5	1622	22	US-09-577-410-7427	Sequence 7427, Ap
3	1232.6	99.5	1656	17	US-09-338-125-2021	Sequence 2021, Ap
4	1232.6	99.5	1656	25	US-09-652-126-9702	Sequence 9702, Ap
5	1232.6	99.5	1656	25	US-09-652-816-9327	Sequence 9327, Ap
6	1232.6	99.5	1656	27	US-09-658-010-14283	Sequence 14283, A
7	1232.6	99.5	1656	27	US-09-698-012-8607	Sequence 8607, Ap
8	1232.6	99.5	1656	28	US-09-710-281-5043	Sequence 5043, Ap
9	1232.6	99.5	1656	28	US-09-716-920-1274	Sequence 1274, Ap
10	1232.6	99.5	1656	29	US-09-726-175-2581	Sequence 2581, Ap
11	1232.6	99.5	1656	29	US-09-726-787-2755	Sequence 2755, Ap
12	1232.6	99.5	1656	29	US-09-726-790-2021	Sequence 2021, Ap
13	1232.6	99.5	1755	30	US-09-732-630-4186	Sequence 4186, Ap
14	1232.6	99.5	1755	30	US-09-770-173-2339	Sequence 2339, Ap
15	1232.6	99.5	2057	30	US-09-760-475-983	Sequence 983, Ap
16	1232.6	99.5	2810	1	PCT-US00-26524B-1503	Sequence 1903, Ap
17	1232.6	99.5	2812	32	PCT-US00-05988-654	Sequence 654, Ap
18	1232.6	99.5	2812	32	US-09-925-300-654	Sequence 654, Ap
19	1231	99.4	1571	56	US-60-239-841-86	Sequence 86, Appl
20	1221.6	98.5	1836	60	US-60-278-258-389	Sequence 389, Appl
21	1220	96.9	1808	17	US-60-164-285-7559	Sequence 7559, Ap
22	1200	96.9	1808	17	US-09-359-922-639	Sequence 639, Appl
23	1200	96.9	1808	17	US-09-359-922-639	Sequence 639, Appl
24	1199.6	96.8	1828	16	US-09-340-623-13594	Sequence 13594, A
25	1199.6	96.8	1828	16	US-09-340-623-13594	Sequence 13594, A
26	1199.6	96.8	1828	31	US-09-898-888-13594	Sequence 13594, A
27	1199.6	96.8	1828	31	US-09-898-888-13594	Sequence 13594, A
28	1093	88.2	164833	54	US-60-216-770-49	Sequence 49, Appl
29	1093	88.2	164833	54	US-60-216-770-49	Sequence 49, Appl
30	1093	88.2	164833	54	US-60-216-770-49	Sequence 49, Appl
31	1058	85.4	191307	54	US-60-243-468-420	Sequence 420, Appl
32	1058	85.4	191307	54	US-60-243-468-420	Sequence 420, Appl
33	1058	85.4	191307	54	US-60-212-664-168	Sequence 168, Appl
34	1058	85.4	191307	54	US-60-212-664-168	Sequence 168, Appl
35	1058	85.4	191307	54	US-60-207-216-1442	Sequence 1442, Appl
36	1058	85.4	191307	54	US-60-207-216-1442	Sequence 1442, Appl
37	1058	85.4	191307	54	US-60-207-216-1442	Sequence 1442, Appl
38	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
39	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
40	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
41	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
42	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
43	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
44	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
45	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
46	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
47	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
48	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
49	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
50	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
51	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
52	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
53	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
54	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
55	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
56	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
57	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
58	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
59	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
60	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
61	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
62	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
63	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
64	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl
65	1058	85.4	191307	54	US-60-212-664-284	Sequence 284, Appl

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C	42	730	58.9	3017	51	US-60-185-362-59	Sequence 59, Appl
C	43	730	58.9	3017	51	US-60-185-362-126	Sequence 126, App
44	723.8	58.4	953	57	US-60-243-468-1942	Sequence 1942, Ap	
45	718.4	58.0	42519	54	US-60-216-770-149	Sequence 149, App	

## ALIGNMENTS

## RESULT 1

US-09-775-693-1

Sequence 1, Application US/09775693

GENERAL INFORMATION:

APPLICANT: Clark, Mike

APPLICANT: Holsberg, Frederick Wayne

APPLICANT: Ensor, Charles Mark

TITLE OF INVENTION: Methods For Predicting Susceptibility Of Patients To Arginine Def

FILE REFERENCE: Therapy

CURRENT APPLICATION NUMBER: US/09/775,693

CURRENT FILING DATE: 2001-02-02

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 1239

TYPE: DNA

ORGANISM: Homo sapiens

US-09-775-693-1

Query Match 100.0%; Score 1239; DB 30; Length 1239;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-295;  
 Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgttcagcaaggctccgtgtctctgtgctacagtgccgctggagacactgtgcatc 60
Db 1 atgttcagcaaggctccgtgtctctgtgctacagtgccgctggagacactgtgcatc 60
QY 61 ctctgtgtgtgtgaaggaagcaagcgtatgaagtcattgtccattctgtgccaacattgtgcag 120
Db 61 ctctgtgtgtgtgaaggaagcaagcgtatgaagtcattgtccattctgtgccaacattgtgcag 120
QY 121 aaggaagacttcgaggaagcaggaagaagcaactgaagcttggggcccaaaagtgctc 180
Db 121 aaggaagacttcgaggaagcaggaagaagcaactgaagcttggggcccaaaagtgctc 180
QY 181 attgaagatgtcagcaaggaattgtggaagattcattcgtgcccagccatccagtcagc 240
Db 181 attgaagatgtcagcaaggaattgtggaagattcattcgtgcccagccatccagtcagc 240
QY 241 gcaactgatagagccgctaccctctgtggaactctctgtgcaagccctgtcatcgcgcgc 300
Db 241 gcaactgatagagccgctaccctctgtggaactctctgtgcaagccctgtcatcgcgcgc 300
QY 301 aaacaagtgaatcgcgcagcggaaggggccaagtatgtgtccacagcgccacagga 360
Db 301 aaacaagtgaatcgcgcagcggaaggggccaagtatgtgtccacagcgccacagga 360
QY 361 aaggggaacatcaggtccggttggtgctgaagctgtactactactggtcccccagataag 420
Db 361 aaggggaacatcaggtccggttggtgctgaagctgtactactactggtcccccagataag 420
QY 421 gtcatgtctccctggagatgcctgaattctacaacccgttcaaggcccgcaatgacctg 480
Db 421 gtcatgtctccctggagatgcctgaattctacaacccgttcaaggcccgcaatgacctg 480
QY 481 atggaatgaagcaagcaacagggattccatcccggttactcccaagaagaccgttgagc 540
Db 481 atggaatgaagcaagcaacagggattccatcccggttactcccaagaagaccgttgagc 540
QY 541 atggaatgaagcaagcaacagggattccatcccggttactcccaagaagaccgttgagc 600
Db 541 atggaatgaagcaagcaacagggattccatcccggttactcccaagaagaccgttgagc 600

```

```

QY 601 caagcgtccacaggtctctacacgaagaccagagccagccaaagccccaacacccct 660
Db 601 caagcgtccacaggtctctacacgaagaccagagccagccaaagccccaacacccct 660
QY 661 gacattctgagatcaggttcaaaaaaggggtccctgtgaagtgacaacagtcacaagat 720
Db 661 gacattctgagatcaggttcaaaaaaggggtccctgtgaagtgacaacagtcacaagat 720
QY 721 ggcacccacccacagactccctgtgagctcttcaatgaactgaacgaagtccgggcaag 780
Db 721 ggcacccacccacagactccctgtgagctcttcaatgaactgaacgaagtccgggcaag 780
QY 781 catggtctggcgttattgacatcgttgagagaacgccttcatctggaatgaagtcacagat 840
Db 781 catggtctggcgttattgacatcgttgagagaacgccttcatctggaatgaagtcacagat 840
QY 841 atctagagagcccccagagagcaccatccttaccatgctcaatttagacatcgagccttc 900
Db 841 atctagagagcccccagagagcaccatccttaccatgctcaatttagacatcgagccttc 900
QY 901 accatgagccgggaagtgcgcaaatcaaaccaagcctgggcttgaatttctgtgagctg 960
Db 901 accatgagccgggaagtgcgcaaatcaaaccaagcctgggcttgaatttctgtgagctg 960
QY 961 gttataccggtttagcgttagccttagccttagtgaaattgtccgcacatgcacgaag 1020
Db 961 gttataccggtttagcgttagccttagccttagtgaaattgtccgcacatgcacgaag 1020
QY 1021 tcccagagcgaattggaaggaagagtgctgtcgttctctaaaggccaggtgtacatc 1080
Db 1021 tcccagagcgaattggaaggaagagtgctgtcgttctctaaaggccaggtgtacatc 1080
QY 1081 ctgcgcggggaagtcgccactgtctctacaaatgaaagctgtgtgacatgaaagtcgac 1140
Db 1081 ctgcgcggggaagtcgccactgtctctacaaatgaaagctgtgtgacatgaaagtcgac 1140
QY 1141 gttgatctagagcaactgatgcacccgggttcattcaaatcaatctccctcagctgaag 1200
Db 1141 gttgatctagagcaactgatgcacccgggttcattcaaatcaatctccctcagctgaag 1200
QY 1201 gaalatcatcgtctccagaaggaagtcactgtccaatag 1239
Db 1201 gaalatcatcgtctccagaaggaagtcactgtccaatag 1239

```

## RESULT 2

US-09-577-410-7427

Sequence 7427, Application US/09577410

GENERAL INFORMATION:

APPLICANT: Gutierrez-Ramos, Jose-Carlos

APPLICANT: Hodge, Martin

APPLICANT: Kingsbury, Gillian

TITLE OF INVENTION: Nucleic Acid Molecules Derived from

FILE REFERENCE: Human T Helper cell, Bone Marrow, and CD34+ Libraries

CURRENT APPLICATION NUMBER: US/09/577,410

CURRENT FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: US 60/135,632

PRIOR FILING DATE: 1999-05-24

PRIOR APPLICATION NUMBER: US 60/135,633

PRIOR FILING DATE: 1999-05-24

PRIOR APPLICATION NUMBER: US 60/135,616

PRIOR FILING DATE: 1999-05-24

NUMBER OF SEQ ID NOS: 8991

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 7427

LENGTH: 1622

TYPE: DNA

ORGANISM: Homo sapiens

US-09-577-410-7427

Query Match 99.5%; Score 1232.6; DB 22; Length 1622;  
 Best Local Similarity 99.7%; Pred. No. 6e-294;  
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 atgtccagcaaaagctccgtgtgtctgtgcttaacagatggtcggtctggaacacctgtgtc 60  
 91 atgtccagcaaaagctccgtgtgtctgtgcttaacagatggtcggtctggaacacctgtgtc 150  
 61 ctctgt 120  
 151 ctctgt 210  
 121 aaggaagacttcggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 180  
 211 aaggaagacttcggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 270  
 181 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 240  
 271 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 330  
 241 gcaactatgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 300  
 331 gcaactatgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 390  
 301 aacaaagtgtgaaatcgcccaagcggaagcggaagcggaagcggaagcggaagcggaagcg 360  
 391 aacaaagtgtgaaatcgcccaagcggaagcggaagcggaagcggaagcggaagcggaagcg 450  
 361 aaggaagcaatcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 420  
 451 aaggaagcaatcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 510  
 421 gtcaatgtccctgtgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 480  
 511 gtcaatgtccctgtgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 570  
 481 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 540  
 571 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 630  
 541 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 600  
 631 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 690  
 601 caagcgctcccaagctctctacacgaagcaagcaagcaagcaagcaagcaagcaagcaagca 660  
 691 caagcgctcccaagctctctacacgaagcaagcaagcaagcaagcaagcaagcaagcaagca 750  
 661 gcaatctctgagatgagatctcaaaaaggggtccctgtgaaagtgtgaaagtgtgaaagt 720  
 751 gcaatctctgagatgagatctcaaaaaggggtccctgtgaaagtgtgaaagtgtgaaagt 810  
 721 ggaac 780  
 811 ggaac 870  
 781 catgt 840  
 871 catgt 930  
 841 atctacagagaccccaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 900  
 931 atctacagagaccccaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 990  
 901 accatgtacccggaagtgtgcaaaatcaaaatcaaaatcaaaatcaaaatcaaaatcaaaat 960  
 991 accatgtacccggaagtgtgcaaaatcaaaatcaaaatcaaaatcaaaatcaaaatcaaaat 1050  
 961 gtgtataccggtttctgtgcaagccctgtgaaattgttgcgcacatgtgcaagcccaag 1020  
 1051 gtgtataccggtttctgtgcaagccctgtgaaattgttgcgcacatgtgcaagcccaag 1110

1021 tcccaagagcagtggaaggaagtgaggtgtccgtctcctaagggcaggtgtacatc 1080  
 1111 tcccaagagcagtggaaggaagtgaggtgtccgtctcctaagggcaggtgtacatc 1170  
 1081 ctgagcgagagtgcccaagctgtctctacacagagtgagtgagtgagtgagtgagtgag 1140  
 1171 ctgagcgagagtgcccaagctgtctctacacagagtgagtgagtgagtgagtgagtgag 1230  
 1141 ggtgatatgagcgaacatgtatgcaagcggtgtacacacacacacacacacacacacacac 1200  
 1231 ggtgatatgagcgaacatgtatgcaagcggtgtacacacacacacacacacacacacacac 1290  
 1201 gaattcatgtctcctaagcaagtgcaactgtcctaagcaag 1239  
 1291 gaattcatgtctcctaagcaagtgcaactgtcctaagcaag 1329

RESULT 3  
 US-09-338-425-2021  
 ; Sequence 2021, Application US/09338425  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williamson, Mark  
 ; APPLICANT: Shyjan, Andrew W.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
 ; TITLE OF INVENTION: THEREFOR  
 ; FILE REFERENCE: 1600 2021-001  
 ; CURRENT APPLICATION NUMBER: US/09/338,425  
 ; CURRENT FILING DATE: 1999-06-22  
 ; PRIOR APPLICATION NUMBER: 60/166,038  
 ; PRIOR FILING DATE: 1999-11-30  
 ; NUMBER OF SEQ ID NOS: 2814  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 2021  
 ; LENGTH: 1656  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-338-425-2021

Query Match 99.5%; Score 1232.6; DB 17; Length 1656;  
 Best Local Similarity 99.7%; Pred. No. 6.1e-294;  
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 atgtccagcaaaagctccgtgtgtctgtgcttaacagatggtcggtctggaacacctgtgtc 60  
 125 atgtccagcaaaagctccgtgtgtctgtgcttaacagatggtcggtctggaacacctgtgtc 184  
 61 ctctgt 120  
 185 ctctgt 244  
 121 aaggaagacttcggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 180  
 245 aaggaagacttcggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 304  
 181 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 240  
 305 atggaagatgtcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 364  
 241 gcaactatgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 300  
 365 gcaactatgaggaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 424  
 301 aacaaagtgtgaaatcgcccaagcggaagcggaagcggaagcggaagcggaagcggaagcg 360  
 425 aacaaagtgtgaaatcgcccaagcggaagcggaagcggaagcggaagcggaagcggaagcg 484  
 361 aaggaagcaatcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 420  
 485 aaggaagcaatcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 544  
 421 gtcaatgtccctgtgaggaagtgagatctcaaaacgggttcaagggccgcaatgtacgt 480

```

Db 545 gtcatctctccttgagagatgcttgaattcttaacacggttcaaggccgaatgacctg 604
QY 481 atgagatagcaagaagcaaggggttcccatcccggtcactcccaagaacccgtggagc 540
Db 605 atggagtagcaagaagcaaggggttcccatcccggtcactcccaagaacccgtggagc 540
QY 541 atgagatagcaagaagcaaggggttcccatcccggtcactcccaagaacccgtggagc 664
Db 665 atgagatagcaagaagcaaggggttcccatcccggtcactcccaagaacccgtggagc 664
QY 601 caagcgcctccaggtctctacagagcccaaggcccaaggcccaacacccct 660
Db 725 caagcgcctccaggtctctacagagcccaaggcccaaggcccaacacccct 784
QY 661 gacattctcgagatcgagttcaaaaagggttccctgtgaggttgagcccaagcagat 720
Db 785 gacattctcgagatcgagttcaaaaagggttccctgtgaggttgagcccaagcagat 844
QY 721 ggcacacccacccagacctctcttgaggtcttcaatgtaactgaagtcgagcagag 780
Db 845 ggcacacccacccagacctctcttgaggtcttcaatgtaactgaagtcgagcagag 904
QY 781 catggtctggccgtatctgaatcgttgagagccggttcaatgtaactgaagtcgagcag 840
Db 905 catggtctggccgtatctgaatcgttgagagccggttcaatgtaactgaagtcgagcag 964
QY 841 atctacagagcccaagcagcagcaccatcttaccatgctcattgaacacagccctc 900
Db 965 atctacagagcccaagcagcagcaccatcttaccatgctcattgaacacagccctc 1024
QY 901 accatgagaccgggaagtgagcaaaatcaaaacaggccctggaggttcaaatgtgagctg 960
Db 1025 accatgagaccgggaagtgagcaaaatcaaaacaggccctggaggttcaaatgtgagctg 1084
QY 961 gtgtacacgggttctgaagctagccctgagtgatgttctgcgcacatgacgcgaag 1020
Db 1085 gtgtacacgggttctgaagctagccctgagtgatgttctgcgcacatgacgcgaag 1144
QY 1021 tcccaagagcagtgaggaagggaaagtgcaggtgtcgtcccaaggccaggtgtacatc 1080
Db 1145 tcccaagagcagtgaggaagggaaagtgcaggtgtcgtcccaaggccaggtgtacatc 1204
QY 1081 ctcggccggaggttcccaactgtctcttcaaaatgagaggtgtgagcatgagcgtgagc 1140
Db 1205 ctcggccggaggttcccaactgtctcttcaaaatgagaggtgtgagcatgagcgtgagc 1264
QY 1141 ggtgatttgagccaactgacgtgacccgggttcaaaatgagaggtgtgagcatgagcgtgagc 1200
Db 1265 ggtgatttgagccaactgacgtgacccgggttcaaaatgagaggtgtgagcatgagcgtgagc 1324
QY 1201 gaatacatgctctccagagcagagtcactgccaataag 1239
Db 1325 gaatacatgctctccagagcagagtcactgccaataag 1363

```

US-09-652-126-9702

```

Query Match          99.5%; Score 1232.6; DB 25; Length 1656;
Best Local Similarity 99.7%; Pred. No. 6.1e-294;
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 atgttcaacaaaggtcctcgtgtctctgacctacagatggcgccctgagacactcgtgac 60
Db 125 atgttcaacaaaggtcctcgtgtctctgacctacagatggcgccctgagacactcgtgac 184
QY 61 ctctgtgtgctgaaggaagcagagcgtatgacgttattgacctatctggccaacattggcag 120
Db 185 ctctgtgtgctgaaggaagcagagcgtatgacgttattgacctatctggccaacattggcag 244
QY 121 aaggaagacttcgaggaagcagcaggaaggaagcactgaaactgtggcgcaaaagggtgtc 180
Db 245 aaggaagacttcgaggaagcagcaggaaggaagcactgaaactgtggcgcaaaagggtgtc 304
QY 181 atgagagatctcagcagaggtgtgtgtgagaggttactatctggcgccatccagtcagc 240
Db 305 atgagagatctcagcagaggtgtgtgtgagaggttactatctggcgccatccagtcagc 364
QY 241 gcaactgtatgaggaacccgtactcctctggtgacactcctctgacagccctgacacgc 300
Db 365 gcaactgtatgaggaacccgtactcctctggtgacactcctctgacagccctgacacgc 424
QY 301 aaacaaagtgaatcgcagcagcggaggggagcgaagtatgttccacagcgcacagga 360
Db 425 aaacaaagtgaatcgcagcagcggaggggagcgaagtatgttccacagcgcacagga 484
QY 361 aaggggaagacatcaggttccggtttgagctcagctgactaacgtgacggcccccagataag 420
Db 485 aaggggaagacatcaggttccggtttgagctcagctgactaacgtgacggcccccagataag 544
QY 421 gtcatgtcctccttgagagatcctcgtatcttaacaaacgggttcaaggccgcaatgacctg 480
Db 545 gtcatgtcctccttgagagatcctcgtatcttaacaaacgggttcaaggccgcaatgacctg 604
QY 481 atgagatcgcgaagaacacagcggatctccatcccggttcaactcccaaggacccgtggagc 540
Db 605 atgagatcgcgaagaacacagcggatctccatcccggttcaactcccaaggacccgtggagc 664
QY 541 atgagatgagaacccatcagacatcagctacgaggtgagatctcttgagaaaccacaagac 600
Db 665 atgagatgagaacccatcagacatcagctacgaggtgagatctcttgagaaaccacaagac 724
QY 601 caagcgcctccaggtctcttcaacgaagaccaggaaccagcccaagcccccacacccct 660
Db 725 caagcgcctccaggtctcttcaacgaagaccaggaaccagcccaagcccccacacccct 784
QY 661 gacattctcgagatcgagttcaaaaagggttccctgtgaggttgagccagtcgaagat 720
Db 785 gacattctcgagatcgagttcaaaaagggttccctgtgaggttgagccagtcgaagat 844
QY 721 ggcacacccacccagacctctcttgaggtcttcaatgtaactgaagtcgagcagag 780
Db 845 ggcacacccacccagacctctcttgaggtcttcaatgtaactgaagtcgagcagag 904
QY 781 catggtctggccgtatctgaatcgttgagagccggttcaatgtaactgaagtcgagcag 840
Db 905 catggtctggccgtatctgaatcgttgagagccggttcaatgtaactgaagtcgagcag 964
QY 841 atctacagagcccaagcagcagcaccatcttaccatgctcattgaacacagccctc 900
Db 965 atctacagagcccaagcagcagcaccatcttaccatgctcattgaacacagccctc 1024
QY 901 accatgagaccgggaagtgagcaaaatcaaaacaggccctggaggttcaaatgtgagctg 960
Db 1025 accatgagaccgggaagtgagcaaaatcaaaacaggccctggaggttcaaatgtgagctg 1084
QY 961 gtgtacacgggttctgaagctagccctgagtgatgttctgcgcacatgacgcgaag 1020

```

```

RESULT 4
US-09-652-126-9702
: Sequence 9702, Application US/09652126
: GENERAL INFORMATION:
: APPLICANT: Shvjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: FILE REFERENCE: 1600,1185-001
: CURRENT APPLICATION NUMBER: US/09/652,126
: PRIOR FILING DATE: 2000-08-30
: PRIOR APPLICATION NUMBER: 60/151,132
: NUMBER OF SEQ ID NOS: 10051
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9702
: LENGTH: 1656
: TYPE: DNA
: ORGANISM: Homo sapiens

```



Db 1085 gtgtatccgttctcgtgcaagccctgagtgatatttgctccgcaactgcatcgccaag 1144  
QY 1021 tcccaagagcgaatggaaggaagtgcaagtgctgctcctcaaggccaggggtatc 1080  
Db 1145 tcccaagagcgaatggaaggaagtgcaagtgctgctcctcaaggccaggggtatc 1204  
QY 1081 ctgcgcggaggtcccaactgctctctacatgaagagctgtgagcatgacgtgcaag 1140  
Db 1205 ctgcgcggaggtcccaactgctctctacatgaagagctgtgagcatgacgtgcaag 1264  
QY 1141 ggtgattatgagccaactgcatgcaacgggttcaatcaatcaatccctcaggtgaa 1200  
Db 1265 ggtgattatgagccaactgcatgcaacgggttcaatcaatcaatccctcaggtgaa 1324  
QY 1201 gaatatcatgcttcccgagagcaagtgcatgccaataag 1239  
Db 1325 gaatatcatgcttcccgagagcaagtgcatgccaataag 1363

RESULT 5  
US-09-652-816-9327  
; Sequence 9327, Application US/09652816  
; GENERAL INFORMATION:  
; APPLICANT: Gutierrez-Ramos, Jose-Carlos  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.1177-001  
; CURRENT APPLICATION NUMBER: US/09/652.816  
; CURRENT FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: 60/152.111  
; PRIOR FILING DATE: 1999-08-31  
; NUMBER OF SEQ ID NOS: 9647  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9327  
; LENGTH: 1656  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-652-816-9327

Query Match 99.5%; Score 1232.6; DB 25; Length 1656;  
Best Local Similarity 99.7%; Pred. No. 6.1e-294;  
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 atgtcagaagaagctcgtgtgtctgtgctacagtgagcgctgagacactgtgcatc 60  
Db 125 atgtcagaagaagctcgtgtgtctgtgctacagtgagcgctgagacactgtgcatc 184  
QY 61 ctctgtgtgtctgaagaaagcgtatgacgtcatctgctctacatgcaacatgtgcaag 120  
Db 185 ctctgtgtgtctgaagaaagcgtatgacgtcatctgctctacatgcaacatgtgcaag 244  
QY 121 aaggaagactcgaagaaagcgaagaaagcgaagcgtgagcgtgggagcgaagagtgctc 180  
Db 245 aaggaagactcgaagaaagcgaagaaagcgaagcgtgagcgtgggagcgaagagtgctc 304  
QY 181 atggaagactcgaagaaagcgtgagaaagtgatcatctgtgcaagcgtcgtgcaagc 240  
Db 305 atggaagactcgaagaaagcgtgagaaagtgatcatctgtgcaagcgtcgtgcaagc 364  
QY 241 gcactgtatgagagcgtctactctctgtggaactctctgtgcaagcgtcgtgcaagc 300  
Db 365 gcactgtatgagagcgtctactctctgtggaactctctgtgcaagcgtcgtgcaagc 424  
QY 301 aagaagatggaatgccaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 360  
Db 425 aagaagatggaatgccaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 484  
QY 361 aaggggaacatcaggtccggttgaagcgtcagcgtgctactcaactggtcccccagataag 420  
Db 485 aaggggaacatcaggtccggttgaagcgtcagcgtgctactcaactggtcccccagataag 544  
QY 421 gtcatgtctcctgtgagagtgctgtaattctacaacgggttcaagcgtcgaatgactgt 480

Db 545 gtcatgtctcctgtgagagtgctgtaattctacaacgggttcaagcgtcgaatgactgt 604  
QY 481 atggaatagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 540  
Db 605 atggaatagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 664  
QY 541 atggaatagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 600  
Db 665 atggaatagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 724  
QY 601 caagcgtcctcaggtgtctctacagaaagcgaagcgaagcgaagcgaagcgaagcgaagc 660  
Db 725 caagcgtcctcaggtgtctctacagaaagcgaagcgaagcgaagcgaagcgaagcgaagc 784  
QY 661 gacatctcagagtcaggttcaaaagaggtccctgtgaagtgagcgaagcgaagcgaagc 720  
Db 785 gacatctcagagtcaggttcaaaagaggtccctgtgaagtgagcgaagcgaagcgaagc 844  
QY 721 ggcac 780  
Db 845 ggcac 904  
QY 781 catggtgtggtcgtatgtgcatcgtggaagcgaagcgaagcgaagcgaagcgaagcgaagc 840  
Db 905 catggtgtggtcgtatgtgcatcgtggaagcgaagcgaagcgaagcgaagcgaagcgaagc 964  
QY 841 atctagagaccccaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 900  
Db 965 atctagagaccccaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 1024  
QY 901 accatgacccggaagtgctgcaaaatcaaaagcgtggtggtgaattgtgtgctgtg 960  
Db 1025 accatgacccggaagtgctgcaaaatcaaaagcgtggtggtgaattgtgtgctgtg 1084  
QY 961 gtgtatccggtttacagcgttagcctgagtggtggtggtggtggtggtggtggtggtggt 1020  
Db 1085 gtgtatccggtttacagcgttagcctgagtggtggtggtggtggtggtggtggtggtggt 1144  
QY 1021 tcccaagagcgaatggaaggaagtgcaagtgctgctcctcaagggccaagtgatcatc 1080  
Db 1145 tcccaagagcgaatggaaggaagtgcaagtgctgctcctcaagggccaagtgatcatc 1204  
QY 1081 ctgcgcggaggtcccaactgctctctacatgaagagctgtgagcatgacgtgcaag 1140  
Db 1205 ctgcgcggaggtcccaactgctctctacatgaagagctgtgagcatgacgtgcaag 1264  
QY 1141 ggtgattatgagccaactgcatgcaacgggttcaatcaatcaatccctcaggtgaa 1200  
Db 1265 ggtgattatgagccaactgcatgcaacgggttcaatcaatcaatccctcaggtgaa 1324  
QY 1201 gaatatcatgcttcccgagagcaagtgcatgccaataag 1239  
Db 1325 gaatatcatgcttcccgagagcaagtgcatgccaataag 1363

RESULT 6  
US-09-698-010-14283  
; Sequence 14283, Application US/09698010  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Mark  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: 1600.2029-001  
; CURRENT APPLICATION NUMBER: US/09/698.010  
; CURRENT FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: 60/162.358  
; PRIOR FILING DATE: 1999-10-29  
; NUMBER OF SEQ ID NOS: 15684  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14283  
; LENGTH: 1656





QY 901 accatggaccgggaagtgcgcaaatcaacaagacctgggttgaaatttgcctgacgtg 960  
 |||||||  
 Db 1025 accatggaccgggaagtgcgcaaatcaacaagacctgggttgaaatttgcctgacgtg 1084  
 QY 961 gtatataccggttaccgacctgagctgtgaattttgcgcgaacctgacgtgcgaag 1020  
 |||||||  
 Db 1085 gtatataccggttaccgacctgagctgtgaattttgcgcgaacctgacgtgcgaag 1144  
 QY 1021 tcccaagagcgagtggaagggaaagtgacagtgctccgtctcaaggcgcaagtgataac 1080  
 |||||||  
 Db 1145 tcccaagagcgagtggaagggaaagtgacagtgctccgtctcaaggcgcaagtgataac 1204  
 QY 1081 ctggcgccgggaagtcgccacctgtctctctcaatagagagctgtgtgacatgacgtgcag 1140  
 |||||||  
 Db 1205 ctggcgccgggaagtcgccacctgtctctctcaatagagagctgtgtgacatgacgtgcag 1264  
 QY 1141 ggtgataatgagcaactatgcaccgggttcataatcaaatccctccgctgcagag 1200  
 |||||||  
 Db 1265 ggtgataatgagcaactatgcaccgggttcataatcaaatccctccgctgcagag 1324  
 QY 1201 gaataatcatcgtctccagagcaaggtcactgtccaaatag 1239  
 |||||||  
 Db 1325 gaataatcatcgtctccagagcaaggtcactgtccaaatag 1363

## RESULT 9

US-09-716-920-1274  
 ; Sequence 1274, Application US/09716920  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lloyd, Clare M.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
 ; FILE REFERENCE: 1600, 2041-001  
 ; CURRENT APPLICATION NUMBER: US/09/716, 920  
 ; CURRENT FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: 60/166,507  
 ; PRIOR FILING DATE: 1999-11-19  
 ; NUMBER OF SEQ ID NOS: 1641  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1274  
 ; LENGTH: 1656  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-716-920-1274

Query Match 99.5%; Score 1232.6; DB 28; Length 1656;  
 Best Local Similarity 99.7%; Pred. No. 6,1e-294;  
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagcaaaagctccgtgttctctgacctagctggcgctggaacctgtgacatc 60  
 |||||||  
 Db 125 atgtccagcaaaagctccgtgttctctgacctagctggcgctggaacctgtgacatc 184  
 QY 61 ctctgttgcgtgaaggaacaaagctatgacttcaatctctgacatctggccaatctggcag 120  
 |||||||  
 Db 185 ctctgttgcgtgaaggaacaaagctatgacttcaatctctgacatctggccaatctggcag 244  
 QY 121 aaggaagacttcgaggaagcagaagaaggaagctggaagcttggggcacaagaagtgttc 180  
 |||||||  
 Db 245 aaggaagacttcgaggaagcagaagaaggaagctggaagcttggggcacaagaagtgttc 304  
 QY 181 attagagatgctcagcaaggagttgtgagagatctatctcggcgccatccagtcacgc 240  
 |||||||  
 Db 305 attagagatgctcagcaaggagttgtgagagatctatctcggcgccatccagtcacgc 364  
 QY 241 gcaatgatatgagacccactactcctctggaacctctctgcagggccctcatgcgcgcg 300  
 |||||||  
 Db 365 gcaatgatatgagacccactactcctctggaacctctctgcagggccctcatgcgcgcg 424  
 QY 301 aaaaagaatggaatcgcgcgcagcgggaaggcccaagatgtgtccacggcgccacaaga 360  
 |||||||  
 Db 425 aaaaagaatggaatcgcgcgcagcgggaaggcccaagatgtgtccacggcgccacaaga 484

QY 361 aagggaacagatcaagatcgcgtttgagctcagctgtactactactggtcccccagataag 420  
 |||||||  
 Db 465 aagggaacagatcaagatcgcgtttgagctcagctgtactactactggtcccccagataag 544  
 QY 421 gtcatctctccctggaagagatccctgaattctcaacaacgggttcgaaggccgaatgacctg 480  
 |||||||  
 Db 545 gtcatctctccctggaagagatccctgaattctcaacaacgggttcgaaggccgaatgacctg 604  
 QY 481 atggagtcgcgaagaacacacgggattcccatcccggttcaactcccaagaaacccgtggagc 540  
 |||||||  
 Db 605 atggagtcgcgaagaacacacgggattcccatcccggttcaactcccaagaaacccgtggagc 664  
 QY 541 atggatggaacctcatgacataagctacagagctgtgaattcctcgggaaccccaagaaac 600  
 |||||||  
 Db 665 atggatggaacctcatgacataagctacagagctgtgaattcctcgggaaccccaagaaac 724  
 QY 601 caagcgctcccaaggctcttacaagaagaccagaagcccaagcccaacacccct 660  
 |||||||  
 Db 725 caagcgctcccaaggctcttacaagaagaccagaagcccaagcccaacacccct 784  
 QY 661 gacattctcgagatcaggtttcaaaaagggtccctgtgaagggtgacaaagtcacagat 720  
 |||||||  
 Db 785 gacattctcgagatcaggtttcaaaaagggtccctgtgaagggtgacaaagtcacagat 844  
 QY 721 ggcacacacacacacagacctctctgtgagctcttatagtactgaagagtcggcggaag 780  
 |||||||  
 Db 845 ggcacacacacacacagacctctctgtgagctcttatagtactgaagagtcggcggaag 904  
 QY 781 catggcgtgggcgttatgtacatcgttgagaacgcgttcatcttgaaatgaatcccgaggt 840  
 |||||||  
 Db 905 catggcgtgggcgttatgtacatcgttgagaacgcgttcatcttgaaatgaatcccgaggt 964  
 QY 841 atctacagagacccagagagacacacatcttaccatgtcatttagacatcgagagcttc 900  
 |||||||  
 Db 965 atctacagagacccagagagacacacatcttaccatgtcatttagacatcgagagcttc 1024  
 QY 901 accatggaccgggaagtgcgcaaatcaacaagacctgggttgaaatttgcctgacgtg 960  
 |||||||  
 Db 1025 accatggaccgggaagtgcgcaaatcaacaagacctgggttgaaatttgcctgacgtg 1084  
 QY 961 gtatataccggttaccgacctgagctgtgaattttgcgcgaacctgacgtgcgaag 1020  
 |||||||  
 Db 1085 gtatataccggttaccgacctgagctgtgaattttgcgcgaacctgacgtgcgaag 1144  
 QY 1021 tcccaagagcgagtggaagggaaagtgacagtgctccgtctcaaggcgcaagtgataac 1080  
 |||||||  
 Db 1145 tcccaagagcgagtggaagggaaagtgacagtgctccgtctcaaggcgcaagtgataac 1204  
 QY 1081 ctggcgccgggaagtcgccacctgtctctctcaatagagagctgtgtgacatgacgtgcag 1140  
 |||||||  
 Db 1205 ctggcgccgggaagtcgccacctgtctctctcaatagagagctgtgtgacatgacgtgcag 1264  
 QY 1141 ggtgataatgagcaactatgcaccgggttcataatcaaatccctccgctgcagag 1200  
 |||||||  
 Db 1265 ggtgataatgagcaactatgcaccgggttcataatcaaatccctccgctgcagag 1324  
 QY 1201 gaataatcatcgtctccagagcaaggtcactgtccaaatag 1239  
 |||||||  
 Db 1325 gaataatcatcgtctccagagcaaggtcactgtccaaatag 1363

## RESULT 10

US-09-726-175-2581  
 ; Sequence 2581, Application US/09726175  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Geating, David P.  
 ; APPLICANT: Fraser, Christopher C.  
 ; APPLICANT: Donovan, Michael J.  
 ; APPLICANT: Holtzman, Douglas A.  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
 ; FILE REFERENCE: 1600, 2054-001

```

: CURRENT APPLICATION NUMBER: US/09/726,175
:
: CURRENT FILING DATE: 2000-11-29
:
: PRIOR APPLICATION NUMBER: 60/167,859
:
: PRIOR FILING DATE: 1999-11-29
:
: NUMBER OF SEO ID NOS: 3770
:
: SOFTWARE: FastSEO for Windows Version 4.0
:
: SEO ID NO 2581
:
: LENGTH: 1656
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
US-09-726-175-2581

```

Query Match	99.58%	Score 1232.6	DB 29	Length 1656
Best Local Similarity	99.78%	Pred. No. 6.1e-294		
Matches 1235	Conservative	0	Mismatches 4	Indels 0
				Gaps 0

OY	1	atgtccacgcaaaagcctccgtagttctctcttgcccttaacagctgagcgcgcctcttggaacacccctctgtacac	50
Db	125	atgtccacgcaaaagcctccgtagttctctcttgcccttaacagctgagcgcgcctcttggaacacccctctgtacac	184
OY	61	ctcgtgtgtgcctgaaagaaacagcgtatgaacgtatcgctatctctgtgccaacatttgccag	120
Db	185	ctcgtgtgtgcctgaaagaaagcgtatgaacgtatcgctatctctgtgccaacatttgccag	244
OY	121	aaggaagaacattccgaggaagcccaagaagaagcgaactgaagctctgaggcccaaaagtgctc	180
Db	245	aaggaagaacattccgaggaagcccaagaagaagcgaactgaagctctgaggcccaaaagtgctc	304
OY	181	atttgaggaatgtccagccagaggaattgttgagagagttcaacttgccagccatccagttccag	240
Db	305	atttgaggaatgtccagccagaggaattgttgagagagttcaacttgccagccatccagttccag	364
OY	241	gaacgtatctgaggaacgcgtctacatccctctcttgccagcccttgccatccgcgcgc	300
Db	365	gaacgtatctgaggaacgcgtctacatccctctcttgccagcccttgccatccgcgcgc	424
OY	301	aaacaagaatgagaaatctgcacacagcggagagggcccaagatgtgtccacagctgcaccaagaa	360
Db	425	aaacaagaatgagaaatctgcacacagcggagagggcccaagatgtgtccacagctgcaccaagaa	484
OY	361	aaagggaacagatccaggtcccggtttgagctcagctgtctactcaactggtgcccccaagataag	420
Db	485	aaagggaacagatccaggtcccggtttgagctcagctgtctactcaactggtgcccccaagataag	544
OY	421	gtcatctgtccctctgaggaagatgcctgaattcttcaacaacgggttcaagagccgcgaattgaacct	480
Db	545	gtcatctgtccctctgaggaagatgcctgaattcttcaacaacgggttcaagagccgcgaattgaacct	604
OY	481	atgagtaagcaagaacaacacaggaattcccaatcccggttcacttcccaagaaccccgtagagc	540
Db	605	atgagtaagcaagaacaacacaggaattcccaatcccggttcacttcccaagaaccccgtagagc	664
OY	541	atgagtaagaaacctcaatgcatacactcaagctacgaagcttgaatctccttgaggaaccccaagaaac	600
Db	665	atgagtaagaaacctcaatgcatacactcaagctacgaagcttgaatctccttgaggaaccccaagaaac	724
OY	601	caagcgccctccaggtctcttcaacacggaagcccaagcccaagcccaagccccaacacccct	660
Db	725	caagcgccctccaggtctcttcaacacggaagcccaagcccaagcccaagccccaacacccct	784
OY	661	gaacatctctgaatctcgaagtctcaaaaaaggggtccctctgtgaagtggaaccaagtcgaagat	720
Db	785	gaacatctctgaatctcgaagtctcaaaaaaggggtccctctgtgaagtggaaccaagtcgaagat	844
OY	721	ggcaccaccaaccaagaacacctcttggaagctcttcatatgtaacctggaacgaagtctggcgggaag	780
Db	845	ggcaccaccaaccaagaacacctcttggaagctcttcatatgtaacctggaacgaagtctggcgggaag	904
OY	781	catggcgttggccggttatgacatctgttggaacacgcttcatattgaattaaatcccgaggt	840
Db	905	catggcgttggccggttatgacatctgttggaacacgcttcatattgaattaaatcccgaggt	964

QY	841	atcaagagacccacgaagacacatcccttaccatgtccattttagacatgagccctc	900
Db	965	atccaagacacccacgaagacacatcccttaccatgtccattttagacatgagccctc	1024
QY	901	accatggaacccggagatgtgcgaataacaagaagcctgggctgtaaatgtctgagctg	960
Db	1025	accatggaacccggagatgtgcgaataacaagaagcctgggctgtaaatgtctgagctg	1084
QY	961	gtgtataccggtttaaagcctagcccttgagtggtgaattgtccgcacatgtacgtccaa	1020
Db	1085	gtgtataccggtttaaagcctagcccttgagtggtgaattgtccgcacatgtacgtccaa	1144
QY	1021	tcccaagagcgagatgtaagaaaggaaagtgacagtggtccgttcctcaagggccaggtgtacac	1080
Db	1145	tcccaagagcgagatgtaagaaaggaaagtgacagtggtccgttcctcaagggccaggtgtacac	1204
QY	1081	ctcgcgcggagatgcccaactgtctctctacaatgagagcgatgtagcatgaagctgaag	1140
Db	1205	ctcgcgcggagatgcccaactgtctctctacaatgagagcgatgtagcatgaagctgaag	1264
QY	1141	gggtattatgagccaactgtatgccaccgggttataataaatcaattccctcgaagctctgaag	1200
Db	1265	gggtattatgagccaactgtatgccaccgggttataataaatcaattccctcgaagctctgaag	1324
QY	1201	gaatatactgtctccagagcaaggttaactgtccaaataag	1239
Db	1325	gaatatactgtctccagagcaaggttaactgtccaaataag	1363

```

RESULT 11
US-09-726-787-2755
: Sequence 2755, Application US/09726787
: GENERAL INFORMATION:
: APPLICANT: Geating, David P.
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEROPOR
: FILE REFERENCE: 1600-2010-001
: CURRENT APPLICATION NUMBER: US/09/726,787
: CURRENT FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: 60/168,132
: PRIOR FILING DATE: 1999-11-30
: NUMBER OF SEQ ID NOS: 3241
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2755
: LENGTH: 1656
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-726-787-2755

```

Query Match	99.5%;	Score 1232.6;	DB 29;	Length 1656;
Best Local Similarity	99.7%;	Pred. No. 6.1e-294;		
Matches 1235; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY	1	atgttcacgaagaagctctccgttgctcttcgtgcacaaagtcgacgagctctgcgaacactctgcatc	60
Db	125	atgttcacgaagaagctctccgttgctcttcgtgcacaaagtcgacgagctctgcgaacactctgcatc	184
QY	61	ctctcgtctgctcgtaaagaaacaagctcctatgacgttcactctctgcgcgaacatttgccag	120
Db	185	ctctcgtctgctcgtaaagaaacaagctcctatgacgttcactctctgcgcgaacatttgccag	244
QY	121	aaggaagaacctctgaaagaagccagaagaagaagcactgtgaagctcttgaggcccaaaagtgcttc	180
Db	245	aaggaagaacctctgaaagaagccagaagaagaagcactgtgaagctcttgaggcccaaaagtgcttc	304
QY	181	attgaagaatcttaagaagaagagctcttgtaagaagttcattcattctgcgcgcgcattccagtcacg	240
Db	305	attgaagaatcttaagaagaagagctcttgtaagaagttcattcattctgcgcgcgcattccagtcacg	364
QY	241	gcactgtatagaaagccgactcctctctgtgacactctcttgccaaagcccttgatgcgcgc	300

```

; Sequence 2021, Application US/09726790
;
; GENERAL INFORMATION:
;
; APPLICANT: Williamson, Mark
;
; APPLICANT: Shvitan, Andrew W.

```

[illegible]



```

Db      |||||||
905      catggcgtggcgctatgacatcgtggaagacgcttcacatggaagtaagcccgaggt 964
Qy      |||||||
841      atctacgagaccccgagcgacacatccttaccatgctacattagacatcgagccttc 900
Db      |||||||
965      atctcggagaccccgagcgacacatccttaccatgctacattagacatcgagccttc 1024
Qy      |||||||
901      accatlgagccgggaaagtgcgcaaaatacaaaagcctgggcttgaatttgcctgagctg 960
Db      |||||||
1025     accatlgagccgggaaagtgcgcaaaatacaaaagcctgggcttgaatttgcctgagctg 1084
Qy      |||||||
961      ggtgtacaggtttacggccttagccttagtgatgttgcctgcacatgcacccaag 1020
Db      |||||||
1085     ggtgtacaggtttctctgacacagccctcgaagtgtgaatttgcctgcacatgcacccaag 1144
Qy      |||||||
1021     tcccgagagcgagtggaaggaagtgacagtgctcgtctcctcaaggcgccagtgctacac 1080
Db      |||||||
1145     tcccgagagcgagtggaaggaagtgacagtgctcgtctcctcaaggcgccagtgctacac 1204
Qy      |||||||
1081     ctccgagcgagtgcccaactgtctcttcaatgagggagctgtgagcatgagctgacag 1140
Db      |||||||
1205     ctccgagcgagtgcccaactgtctcttcaatgagggagctgtgagcatgagctgacag 1264
Qy      |||||||
1141     ggtgtatagagcaactatgcacacgggttcacatcaaatcctcctcagctgagag 1200
Db      |||||||
1265     ggtgtatagagcaactatgcacacgggttcacatcaaatcctcctcagctgagag 1324
Qy      |||||||
1201     gaatacatcgtctccagagcaagtgctactgccaatag 1239
Db      |||||||
1325     gaatacatcgtctccagagcaagtgctactgccaatag 1363

```

## RESULT 13

```

US-09-732-630-4186
; Sequence 4186, Application US/09732630
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Deeds, James
; APPLICANT: Berger, Allison
; APPLICANT: Zhao, Xunel
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF CERVICAL CANCER
; FILE REFERENCE: MRI-008B
; CURRENT APPLICATION NUMBER: US/09/732,630
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/169,681
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/171,350
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: US 60/189,315
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: US 60/203,791
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: US 60/210,600
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 60/220,114
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 4660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4186
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1,2,1725
; OTHER INFORMATION: n = a,c,g, or t
US-09-732-630-4186

```

```

Query Match      99.5%; Score 1232.6; DB 29; Length 1725;
Best Local Similarity 99.7%; Pred. No. 6,1e-294;

```

```

Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy      1 atctcagcaaaagctccgctgtgtctggtcctacagtgagcgccctgacacccctgtacac 60
Db      132 atgtccagcaaaagctccgctgtgtgtcgtcctacagtgagcgccctgacacccctgtacac 191
Qy      61 ctgctgtgtgtgaaggaacaaagctatgacgttactgtcctactgtgccaatattgccaag 120
Db      192 ctgctgtgtgtgaaggaacaaagctatgacgttactgtcctactgtgccaatattgccaag 251
Qy      121 aagaaagacttcgagagagcgacgaagaagaagcactgaagcttggggccaagaagtgcttc 180
Db      252 aagaaagacttcgagagagcgacgaagaagaagcactgaagcttggggccaagaagtgcttc 311
Qy      181 attagatgtcagcaagagaggttctgtgaggaagttcactgtgcccgcacatccagtcacac 240
Db      312 attagatgtcagcaagagaggttctgtgaggaagttcactgtgcccgcacatccagtcacac 371
Qy      241 gcactgtatgagagccgctacactcctctgggacacccctcttgcagagccctgacatgcgcgc 300
Db      372 gcactgtatgagagccgctacactcctctgggacacccctcttgcagagccctgacatgcgcgc 431
Qy      301 aaaaagtggaagaaatccgcccagcggaaggggccaagtatgtgtcccaagcgccacaga 360
Db      432 aaaaagtggaagaaatccgcccagcggaaggggccaagtatgtgtcccaagcgccacaga 491
Qy      361 aagggaaacgatacagtgctcgttctgagctcagctgactcactcgtgcccacgaataag 420
Db      492 aagggaaacgatacagtgctcgttctgagctcagctgactcactcgtgcccacgaataag 551
Qy      421 gtcattgtcctctgagagagatgtcctgaattctacaacgggttcgaaggcgccgaatgactg 480
Db      552 gtcattgtcctctgagagagatgtcctgaattctacaacgggttcgaaggcgccgaatgactg 611
Qy      481 atggaatgcgcaaaacgaacaggaattccatcccggtcactcccaagaacccgtgagagc 540
Db      612 atggaatgcgcaaaacgaacaggaattccatcccggtcactcccaagaacccgtgagagc 671
Qy      541 atggaatgcgcaaaacgaacaggaattccatcccggtcactcccaagaacccgtgagagc 600
Db      672 atggaatgcgcaaaacgaacaggaattccatcccggtcactcccaagaacccgtgagagc 731
Qy      601 caagcgccctccaggtgtctctacacgagaccccgagcccgccaagccccaacacccct 660
Db      732 caagcgccctccaggtgtctctacacgagaccccgagcccgccaagccccaacacccct 791
Qy      661 gacattctcgagatcaggttcaaaaaggggtccctgtgaaagtgtgaaagtgcaagtgcaagat 720
Db      792 gacattctcgagatcaggttcaaaaaggggtccctgtgaaagtgtgaaagtgcaagtgcaagat 851
Qy      721 ggcacaccccaacgaacacccctctgtgagctcttcattgacttgaagaaatgcgggcaag 780
Db      852 ggcacaccccaacgaacacccctctgtgagctcttcattgacttgaagaaatgcgggcaag 911
Qy      781 catgagtgaggccgtatctacatcgtgagaaacgggttacttgaatgaagtcccgaggt 840
Db      912 catgagtgaggccgtatctacatcgtgagaaacgggttacttgaatgaagtcccgaggt 971
Qy      841 atctacgagaccccgagcgacacatcccttaccatgctcattttagacatcgagagccttc 900
Db      972 atctacgagaccccgagcgacacatcccttaccatgctcattttagacatcgagagccttc 1031
Qy      901 accatlgagccgggaaagtgcgcaaaatacaaaagcctgggcttgaatttgcctgagctg 960
Db      1032 accatlgagccgggaaagtgcgcaaaatacaaaagcctgggcttgaatttgcctgagctg 1091
Qy      961 ggtgtacaggtttacggccttagccttagtgatgttgcctgcacatgcacccaag 1020
Db      1092 ggtgtacaggtttctctgacacagccctcgaagtgtgaatttgcctgcacatgcacccaag 1151
Qy      1021 tcccgagagcgagtggaaggaagtgacagtgctcgtctcctcaaggcgccagtgctacac 1080
Db      1152 tcccgagagcgagtggaaggaagtgacagtgctcgtctcctcaaggcgccagtgctacac 1211

```

QY 1081 ctgagccggagatcccaactgtctctataatgagagctgtgagcatgaacgtgcag 1140  
 Db 1212 ctgagccggagatcccaactgtctctataatgagagctgtgagcatgaacgtgcag 1271  
 QY 1141 ggtgattatgagcaactatgacccgggttcatcaataatccctcagctgaag 1200  
 Db 1272 ggtgattatgagcaactatgacccgggttcatcaataatccctcagctgaag 1331  
 QY 1201 gaatacatcgtctccagagcaagctcactgcgaatag 1239  
 Db 1332 gaatacatcgtctccagagcaagctcactgcgaatag 1370

RESULT 14  
 US-09-770-173-2339

Sequence 2339, Application US/09770173  
 GENERAL INFORMATION:  
 APPLICANT: White, David  
 APPLICANT: Pan, Yang  
 TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
 TITLE OF INVENTION: THEREFOR  
 FILE REFERENCE: 1600.2059-001  
 CURRENT APPLICATION NUMBER: US/09/770,173  
 CURRENT FILING DATE: 2001-01-26  
 PRIOR APPLICATION NUMBER: US 60/178,876  
 PRIOR FILING DATE: 2000-01-28  
 NUMBER OF SEQ ID NOS: 3167  
 SOFTWARE: PastSeq for Windows Version 4.0  
 SEQ ID NO 2339  
 LENGTH: 1755  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-770-173-2339

Query Match 99.5%; Score 1232.6; DB 30; Length 1755;  
 Best Local Similarity 99.7%; Pred. No. 6.2e-294;  
 Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagcaaaagctcgtgtgtctgtgctcactgagtgagcgcttgacacctgtgac 60  
 Db 131 atgtccagcaaaagctcgtgtgtctgtgctcactgagtgagcgcttgacacctgtgac 120  
 QY 61 ctctgtgtctgaagagcaagctatgacgtcatgtccatctgtgccaacatctgtccag 120  
 Db 191 ctctgtgtctgaagagcaagctatgacgtcatgtccatctgtgccaacatctgtccag 250  
 QY 121 aaggaagacttcgaagagcaagcaaggaagcactgaagcttggggccaanaagtgctc 180  
 Db 251 aaggaagacttcgaagagcaagcaaggaagcactgaagcttggggccaanaagtgctc 310  
 QY 181 atcgagatgtcagcagagagcttctgtgagagatcatctcgccgcgcacatccagtcacg 240  
 Db 311 atcgagatgtcagcagagagcttctgtgagagatcatctcgccgcgcacatccagtcacg 370  
 QY 241 gcaactatgagagcagcctactcctctgtgacacctcctctgcaagcgccctgacgcgcg 300  
 Db 371 gcaactatgagagcagcctactcctctgtgacacctcctctgcaagcgccctgacgcgcg 430  
 QY 301 aacaagtggaatcgcagcagcgaggggggccaagtatgttccacggcgccaagga 360  
 Db 431 aacaagtggaatcgcagcagcgaggggggccaagtatgttccacggcgccaagga 490  
 QY 361 aagggagagcagatccagcttctgagctcagctcactcactgcggcccccaagtaag 420  
 Db 491 aagggagagcagatccagcttctgagctcagctcactcactgcggcccccaagtaag 550  
 QY 421 gtcatgtctccttgagagatgctccttaattctacaacgggttcaagggcgcaatgacctg 480  
 Db 551 gtcatgtctccttgagagatgctccttaattctacaacgggttcaagggcgcaatgacctg 610  
 QY 481 atggagtgacgcaaaacagcggaattcccatcccggtcactcccaagaacccgtlgaagc 540

Db 611 atggagtgacgcaaaacagcggaattcccatcccggtcactcccaagaacccgtlgaagc 670  
 QY 541 atggaatgaacacctatgcacatcagctagagcttgaatccctggagaaacccaagaac 600  
 Db 671 atggaatgaacacctatgcacatcagctagagcttgaatccctggagaaacccaagaac 730  
 QY 601 caagcgctccaggttctctacagaaagccagagaccagccaagcccccaacacccct 660  
 Db 731 caagcgctccaggttctctacagaaagccagagaccagccaagcccccaacacccct 790  
 QY 661 gacattctcgagatccaggttcaaaaagggtccctgtgaaagtgaccaagctcaagat 720  
 Db 791 gacattctcgagatccaggttcaaaaagggtccctgtgaaagtgaccaagctcaagat 850  
 QY 721 ggcacccacccagacacctcttgagctctcatgaccttgaagcaagctcgaggcaag 780  
 Db 851 ggcacccacccagacacctcttgagctctcatgaccttgaagcaagctcgaggcaag 910  
 QY 781 catggcgtggccgtatgtgacatcgttgagaaacccgttcaatggaaatgagtcaggat 840  
 Db 911 catggcgtggccgtatgtgacatcgttgagaaacccgttcaatggaaatgagtcaggat 970  
 QY 841 atctacagagccccaagcagcagccatcccttaccatgctcatgtagacatcgagccttc 900  
 Db 971 atctacagagccccaagcagcagccatcccttaccatgctcatgtagacatcgagccttc 1030  
 QY 901 accatgagccgggaagtgcgcaaaatcaaacagagcctgggttgaatcttctgagctg 960  
 Db 1031 accatgagccgggaagtgcgcaaaatcaaacagagcctgggttgaatcttctgagctg 1090  
 QY 961 gtgtataccggttaccgagctagccctgagtgatgttgcgcacatgcacatgcagcag 1020  
 Db 1091 gtgtataccggttaccgagctagccctgagtgatgttgcgcacatgcacatgcagcag 1150  
 QY 1021 tcccaagagcagatgtgaaaggaagatgacagctgtcctcctcaagggcgcaagtgatc 1080  
 Db 1151 tcccaagagcagatgtgaaaggaagatgacagctgtcctcctcaagggcgcaagtgatc 1210  
 QY 1081 ctgagccggagatccccaactgtctctctacaaatgagagctgtgtgacatgaacgtgcag 1140  
 Db 1211 ctgagccggagatccccaactgtctctctacaaatgagagctgtgtgacatgaacgtgcag 1270  
 QY 1141 ggtgattatgagcaactatgacccgggttcatcaataatccctcagctgaag 1200  
 Db 1271 ggtgattatgagcaactatgacccgggttcatcaataatccctcagctgaag 1330  
 QY 1201 gaatacatcgtctccagagcaagctcactgcgaatag 1239  
 Db 1331 gaatacatcgtctccagagcaagctcactgcgaatag 1369

RESULT 15  
 US-09-760-475-983

Sequence 983, Application US/09760475  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT249  
 CURRENT APPLICATION NUMBER: US/09/760,475  
 CURRENT FILING DATE: 2001-01-16  
 Prior application data removed - consult PALM or file wrapper  
 NUMBER OF SEQ ID NOS: 4122  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 983  
 LENGTH: 2057  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (2036)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE



LOCATION: (2038)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-760-475-983

Query Match 99.5%; Score 1232.6; DB 30; Length 2057;  
Best Local Similarity 99.7%; Pred. No. 6,4e-294;  
Matches 1235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgtccagcaaaagctccgtggtctctgctctacagtgccggtccgacactcgtgac 60  
DB 533 atgtccagcaaaagctccgtggtctctgctctacagtgccggtccgacactcgtgac 592  
QY 61 ctctgtggtccgaaagcaaaagctctacagtgctctacactctgccaacatctgcccag 120  
DB 593 ctctgtggtccgaaagcaaaagctctacagtgctctacactctgccaacatctgcccag 652  
QY 121 aagaaagactctcgaagcaagcagaagcaagcaactgaagcttggggccaaaagtgctc 180  
DB 653 aagaaagactctcgaagcaagcagaagcaagcaactgaagcttggggccaaaagtgctc 712  
QY 181 attgagatgtcagcagagagattgtgagagattcattctgcccggccatctcagc 240  
DB 713 attgagatgtcagcagagagattgtgagagattcattctgcccggccatctcagc 772  
QY 241 gcaactgtatgagagacgtctacactctctggacactctcttgcagagccctgcatcgccgc 300  
DB 773 gcaactgtatgagagacgtctacactctctggacactctcttgcagagccctgcatcgccgc 832  
QY 301 aaacaagctggaatctgcacacagcgaggggagcgcaagatgtctgccacgagccacaagga 360  
DB 833 aaacaagctggaatctgcacacagcgaggggagcgcaagatgtctgccacgagccacaagga 892  
QY 361 aaggggaacgatacagatccggtttgagctcagctgactactgagcccccagataag 420  
DB 893 aaggggaacgatacagatccggtttgagctcagctgactactgagcccccagataag 952  
QY 421 gtcattgtctccctgagagatgctctaatcttcaaacacggttcaagggccgcaatgacctg 480  
DB 953 gtcattgtctccctgagagatgctctaatcttcaaacacggttcaagggccgcaatgacctg 1012  
QY 481 atggaagctacgcaaaacaaacagggatctccatcccggtcactccccaagaacccgtgagc 540  
DB 1013 atggaagctacgcaaaacaaacagggatctccatcccggtcactccccaagaacccgtgagc 1072  
QY 541 atggaagctacgcaaaacaaacagggatctccatcccggtcactccccaagaacccgtgagc 600  
DB 1073 atggaagctacgcaaaacaaacagggatctccatcccggtcactccccaagaacccgtgagc 1132  
QY 601 caagcgctccaggtctctacacgaagacccagacccagccaagcccccaacacccct 660  
DB 1133 caagcgctccaggtctctacacgaagacccagacccagccaagcccccaacacccct 1192  
QY 661 gacattctcgaagatcagatctcaaaaaaagggctccgttgaagtgacaacgtcaaggt 720  
DB 1193 gacattctcgaagatcagatctcaaaaaaagggctccgttgaagtgacaacgtcaaggt 1252  
QY 721 ggcacacccaacagacccctcttgagctcttcatgtacctaagcaagtcgcgggcaag 780  
DB 1253 ggcacacccaacagacccctcttgagctcttcatgtacctaagcaagtcgcgggcaag 1312  
QY 781 catggtgtggtccgttatctgacatcgtggaagacgcttcaattggaatgaagtcgccgaggt 840  
DB 1313 catggtgtggtccgttatctgacatcgtggaagacgcttcaattggaatgaagtcgccgaggt 1372  
QY 841 atttaagaagaccccaagcaagacacactcttaccatgctcattttagacatcgagccctc 900  
DB 1373 atttaagaagaccccaagcaagacacactcttaccatgctcattttagacatcgagccctc 1432  
QY 901 accattgacccgggaagtgcgcaaatcaaaacaaagccttggttgaaatttgctgaagctg 960  
DB 1433 accattgacccgggaagtgcgcaaatcaaaacaaagccttggttgaaatttgctgaagctg 1492

QY 961 gtgtataccggttttaacgacctagccctgagtgtaattgtctgacactgtacgtccaaag 1020  
DB 1493 gtgtataccggttttaacgacctagccctgagtgtaattgtctgacactgtacgtccaaag 1552  
QY 1021 tcccaagagcagtgagaaagtgacaggtgtccgtctcctcaagggccaagtgtacalc 1080  
DB 1553 tcccaagagcagtgagaaagtgacaggtgtccgtctcctcaagggccaagtgtacalc 1612  
QY 1081 ctccgagcaggttcccaactgtctctctacaaatgagagctgtgagcaatgaacgtgacag 1140  
DB 1613 ctccgagcaggttcccaactgtctctctacaaatgagagctgtgagcaatgaacgtgacag 1672  
QY 1141 ggtgattatgagccaactatgtccacacgggtttcatcaacaatcaattccctcaagctgag 1200  
DB 1673 ggtgattatgagccaactatgtccacacgggtttcatcaacaatcaattccctcaagctgag 1732  
QY 1201 gaatatcatgctctccagagcaaggtcactgtgccaatag 1239  
DB 1733 gaatatcatgctctccagagcaaggtcactgtgccaatag 1771

Search completed: February 12, 2002, 14:27:58  
Job time: 6091 sec

Wed Feb 13 07:36:04 2002

us-09-775-693-1.rmpm

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2002, 13:10:07 ; Search time 4918.34 Seconds  
(without alignments)  
598.044 Million cell updates/sec

Title: US-09-775-693-1

Perfect score: 1239  
Sequence: 1 atgtcagcaagagcctcgt.....gcaagtcactgccaatag 1239

Scoring table: IDENTITY-NUC  
Gapop 10.0, Gapext 1.0

Searched: 2043436 seqs, 118699438 residues

Total number of hits satisfying chosen parameters: 4086872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/pdata/2/pna/pct\_NEW\_COMB.seq.\*  
2: /cgn2\_6/pdata/2/pna/US06\_NEW\_COMB.seq.\*  
3: /cgn2\_6/pdata/2/pna/US07\_NEW\_COMB.seq.\*  
4: /cgn2\_6/pdata/2/pna/US08\_NEW\_COMB.seq.\*  
5: /cgn2\_6/pdata/2/pna/US09\_NEW\_COMB.seq.\*  
6: /cgn2\_6/pdata/2/pna/US10\_NEW\_COMB.seq.\*  
7: /cgn2\_6/pdata/2/pna/US11\_NEW\_COMB.seq.\*  
8: /cgn2\_6/pdata/2/pna/US12\_NEW\_COMB.seq.\*  
9: /cgn2\_6/pdata/2/pna/US13\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1231	99.4	1571	US-09-981-353-86	Sequence 86, App1
2	1199.6	96.8	1828	US-09-898-888A-13594	Sequence 13594, A
3	1183.8	95.5	2004	US-60-340-187-216	Sequence 216, App
4	1183.8	95.5	2005	US-60-340-187-310	Sequence 310, App
5	742.8	60.0	1871	US-60-340-187-951	Sequence 951, App
6	554.6	44.8	987	US-09-898-888A-22919	Sequence 22919, A
7	503	43.7	543	US-09-998-598-467	Sequence 467, App
8	503	40.6	503	US-09-998-598-702	Sequence 702, App
9	503	40.6	503	US-09-998-598-907	Sequence 907, App
10	503	40.6	503	US-09-998-598-924	Sequence 924, App
11	499.8	40.3	503	US-09-998-598-1365	Sequence 1365, App
12	467.4	37.7	469	US-09-998-598-787	Sequence 787, App
13	457.4	36.9	478	US-09-933-524A-54674	Sequence 54674, A
14	454	36.6	538	US-10-029-386-8869	Sequence 8869, App
15	453.2	36.6	465	US-09-922-340-266	Sequence 266, App
16	453.2	36.6	465	US-09-922-340A-266	Sequence 266, App
17	434.8	35.1	499	US-09-998-598-883	Sequence 883, App
18	432.8	34.9	474	US-09-904-013-19540	Sequence 19540, A
19	432	34.9	452	US-09-933-524A-63203	Sequence 63203, A
20	430	34.7	443	US-09-933-524A-54673	Sequence 54673, A
21	422	34.1	454	US-09-898-888A-22065	Sequence 22065, A
22	421.8	34.0	453	US-09-939-397-38994	Sequence 38994, A
23	405.4	32.7	408	US-09-939-397-1772	Sequence 1772, App
24	404.8	32.7	492	US-10-029-386-22593	Sequence 22593, A

25	403	32.5	412	7	US-09-925-564-28045	Sequence 28045, A
26	398	32.1	398	6	US-09-927-875A-6305	Sequence 6305, App
27	397.8	32.1	426	6	US-09-933-524A-49321	Sequence 49321, A
28	389.8	31.5	435	6	US-09-933-524A-58820	Sequence 58820, A
29	385.8	31.1	505	7	US-09-925-564-6718	Sequence 6718, App
30	384.2	31.0	400	8	US-10-029-386-9059	Sequence 9059, App
31	383.4	30.9	385	7	US-09-939-397-1771	Sequence 1771, App
32	382.2	30.8	403	5	US-09-904-013-16995	Sequence 16995, A
33	379.6	30.6	401	6	US-09-933-524A-92776	Sequence 92776, A
34	376.6	30.4	404	6	US-09-925-564-28304	Sequence 28304, A
35	376	30.3	377	7	US-09-939-397-47	Sequence 47, App1
36	371.4	30.0	459	6	US-09-975-640-11846	Sequence 11846, A
37	367.4	30.0	459	6	US-09-975-640A-11846	Sequence 11846, A
38	367.4	29.7	468	7	US-09-925-564-28044	Sequence 28044, A
39	354.4	28.6	382	7	US-09-925-564-28986	Sequence 28986, A
40	353	28.5	394	7	US-09-925-564-21043	Sequence 21043, A
41	352.2	28.4	359	5	US-09-904-013-29730	Sequence 29730, A
42	351.2	28.3	385	7	US-09-925-564-6780	Sequence 6780, App
43	350	28.2	351	6	US-09-975-640-26735	Sequence 26735, A
44	350	28.2	351	6	US-09-975-640A-26735	Sequence 26735, A
45	349.8	28.2	363	7	US-09-939-397-32843	Sequence 32843, A

## ALIGNMENTS

RESULT 1  
US-09-981-353-86  
Sequence 86, Application US/09981353  
GENERAL INFORMATION:  
APPLICANT: Jones, David A.  
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
FILE REFERENCE: PA-0038 US  
CURRENT APPLICATION NUMBER: US/09/981,353  
CURRENT FILING DATE: 2001-10-11  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 86  
LENGTH: 1571  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID No: 1543330CB1  
US-09-981-353-86

Query Match	99.4%	Score 1231	DB 6	Length 1571
Best local similarity	99.6%	Pred. No. 0		
Matches 1234	Conservative	0	Mismatches	5
			Indels	0
			Gaps	0
QY 1	atgtcagcaagagcctcgtggtctgtgctacacatgagcgccctggaacctcgtgcatc	60		
DB 89	atgtcagcaagagcctcgtggtctgtgctacacatgagcgccctggaacctcgtgcatc	148		
QY 61	ctcgtgtgctgaggaagaaagcgtatgacgtcatctgtgctatctgtgccaatctgtgcca	120		
DB 149	ctcgtgtgctgaggaagaaagcgtatgacgtcatctgtgctatctgtgccaatctgtgcca	208		
QY 121	aaggaagacttgaggaagcgcaggaaggaagcactgaaccttggggccaagaagtgttc	180		
DB 209	aaggaagacttgaggaagcgcaggaaggaagcactgaaccttggggccaagaagtgttc	268		
QY 181	atgaagatctcagcaagaggttctgtgaggaattatctgtgccaatctgtgccaatctgtgcca	240		
DB 269	atgaagatctcagcaagaggttctgtgaggaattatctgtgccaatctgtgccaatctgtgcca	328		
QY 241	gcactgtatgaggaagcgcactcctcctgtggaaccccttggccaagcctgcatgcccgc	300		
DB 329	gcactgtatgaggaagcgcactcctcctgtggaaccccttggccaagcctgcatgcccgc	388		
QY 301	aaacaagtggaatcgcacagcgaggggccaagtatgtgccagggcgccaagga	360		

```

? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: us/09/340,623
? PRIOR FILING DATE: 1999-06-28
? PRIOR APPLICATION NUMBER: US 09/205,070
? PRIOR FILING DATE: 1998-12-03
? NUMBER OF SEQ ID NOS: 45207
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 13594
? LENGTH: 1828
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(1828)
? OTHER INFORMATION: n = A,T,C or G
US-09-898-888A-13594

Query Match          96.88;   Score 1199.6;   DB 6;   Length 1828;
Best Local Similarity 99.48;   Pred. No. 0;
Matches 1235;   Conservative 0;   Mismatches 4;   Indels 3;   Gaps 3

```

Query Match	96.88;	Score 1199.6;	DB 6;	Length 1828;
Best Local Similarity	99.48;	Pred. No. 0;		
Matches 1235;	Conservative	0;	Mismatches 4;	Indels 3;
			Gaps	3;

Db	749	gaattctctcgagatctcgagttccaataaaggggtccctcttgaaaggtgacccaagtccaagat	720
QY	721	ggcaaccaccacacagactcctctggagctctctcaatgtaacctgaagaagtctgcggcgaag	780
Db	809	ggcaccacaccacacagactcctcttgagctctctcaatgtaacctgaagaagtcgcggcgaag	868
QY	781	catggcgtgggcgcatgtaatactgtgagaaaccgctctcaatgtaagtgaatcccgagt	840
Db	869	catggcgtgggcgcatgtaatactgtgagaaaccgctctcaatgtaagtgaatcccgagt	928
QY	841	attctacgagaccacacagcagcagacatcccttaccatgctccatttgaatactgagagcttc	900
Db	929	attctacgagaccacacagcagcagacatcccttaccatgctccatttgaatactgagagcttc	988
QY	901	accatcgagccgggaagtgcaccaaatccaacaagagcctggagcttgaatatctgcagatcg	960
Db	989	accatcgagccgggaagtgcaccaaatccaacaagagcctggagcttgaatatctgcagatcg	1048
QY	961	gtgatatccggtttacgcgacctagccctctggagtgaattgtgcgcgcacatgcacgcgaag	1020
Db	1049	gtgatatccggtttctctgcgcacagccctctggagtgaattgtgcgcgcacatgcacgcgaag	1108
QY	1021	ttcccagagagcggaagtgaagaaggaaatgtcagatgtctcgtccccaaggcgccaggtgtacatc	1080
Db	1109	ttcccagagagcggaagtgaagaaggaaatgtcagatgtctcgtccccaaggcgccaggtgtacatc	1168
QY	1081	ctcgcgcgagagctccccaactgtctctctacataatgaagagctgctgagcatgaaagtgcag	1140
Db	1169	ctcgcgcgagagctccccaactgtctctctacataatgaagagctgctgagcatgaaagtgcag	1228
QY	1141	ggtatattatagccaactgtatgtccacgcgggtttcaataccaatctccctcagctgtgaag	1200
Db	1229	ggtatattatagccaactgtatgtccacgcgggtttcaataccaatctccctcagctgtgaag	1288
QY	1201	gaatatcatgtctccacaagcagaaggtacactgcacgaataag	1237
Db	1289	gaatatcatgtctccacaagcagaaggtacactgcacgaataag	1327

```

RESULT      2
US-09-898-888A-13594
; Sequence 13594, Application US/09898888A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONFIGS OBTAINED FROM VARIOUS CDNA
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: 20411-748CON1
; CURRENT APPLICATION NUMBER: US/09/898,888A

```

[illegible]

```

Oy 778 aagatgacgtgagccgctatgacatcgtgagaaacccgttcattggaatgaatccga 837
    |||||||
Db 856 aagatgacgtgagccgctatgacatcgtgagaaacccgttcattggaatgaatccga 915
Oy 838 ggtatctagagaccacagcagcagcaccatcccttaccatcttattgaacatcagagcc 897
    |||||||
Db 916 ggtatctagagaccacagcagcagcaccatcccttaccatcttattgaacatcagagcc 975
Oy 898 ttaccatgagccgagaaatgctgcaaatcaaaacagcctgggcttgaatctgtcag 957
    |||||||
Db 976 ttaccatgagccgagaaatgctgcaaatcaaaacagcctgggcttgaatctgtcag 1035
Oy 958 ctgtgtatatacgggttaccggtacgctgagtgatgtatcttctccgcccctgacgccc 1017
    |||||||
Db 1036 ctgtgtatatacgggttaccggtacgctgagtgatgtatcttctccgcccctgacgccc 1095
Oy 1018 aagtcacagagagcgagtgagaggaagtgagtgatgtctcctcgaaggccagtgatc 1077
    |||||||
Db 1096 aagtcacagagagcgagtgagaggaagtgagtgatgtctcctcgaaggccagtgatc 1155
Oy 1078 atcctcgccgagagtgcccaactgtctctcaaatgagagcgctgtgtgagcagtgatc 1137
    |||||||
Db 1156 atcctcgccgagagtgcccaactgtctctcaaatgagagcgctgtgtgagcagtgatc 1215
Oy 1138 caggtgtatatacgggttaccggtacgctgagtgatgtatcttctccgcccctgacgccc 1197
    |||||||
Db 1216 caggtgtatatacgggttaccggtacgctgagtgatgtatcttctccgcccctgacgccc 1275
Oy 1198 aaggaatatcatcgtctccagagcaaggttcatctgccaatag 1239
    |||||||
Db 1276 aaggaatatcatcgtctccagagcaaggttcatctgccaatag 1317

```

# RESULT 3

US-60-187-216

Sequence 216, Application US/60340187

GENERAL INFORMATION:

APPLICANT: Wang, Y. Tom

APPLICANT: Wang, Jian-Rui

APPLICANT: Ghosh, Malabika

APPLICANT: Meng, Gezh

APPLICANT: Boyle, Bryan J

APPLICANT: Drmanac, Radoje T

TITLE OF INVENTION: Novel Nucleic Acids and

Polypeptides

FILE REFERENCE: 813

CURRENT APPLICATION NUMBER: US/60/340,187

CURRENT FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: US 09/488,725

PRIOR FILING DATE: 2000-01-21

PRIOR APPLICATION NUMBER: US 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: PCT/US00/35017

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: US 09/491,404

PRIOR FILING DATE: 2000-01-25

PRIOR APPLICATION NUMBER: PCT/US01/02623

PRIOR FILING DATE: 2001-01-25

PRIOR APPLICATION NUMBER: US 09/496,914

PRIOR FILING DATE: 2000-02-03

PRIOR APPLICATION NUMBER: US 09/560,875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: PCT/US01/03800

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 09/515,126

PRIOR FILING DATE: 2000-02-28

PRIOR APPLICATION NUMBER: US 09/577,409

PRIOR FILING DATE: 2000-05-18

PRIOR APPLICATION NUMBER: PCT/US01/04927

PRIOR FILING DATE: 2001-02-26

PRIOR APPLICATION NUMBER: US 09/519,705

```

PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/574,454
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: PCT/US01/04941
PRIOR FILING DATE: 2001-03-05
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: US 09/649,167
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/552,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 09/770,160
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: PCT/US01/08656
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 09/577,408
PRIOR FILING DATE: 2000-05-18
PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 1192
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO: 216
LENGTH: 2004
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (95)...(346)
US-60-340-187-216

```

```

Query Match 95.5%; Score 1183.8; DB 9; Length 2004;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1188; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 1 atctcaagcaaaagctccgtgttctgtctacaaatgagcggccttgacacctctgacatc 60
    |||||||
Db 481 atgtccagcaaaagctccgtgttctgtctacaaatgagcggccttgacacctctgacatc 540
Oy 61 ctctgtgtctgaagaaacaaagctatgacgtcattgtcctatcttgccaaattgtgacag 120
    |||||||
Db 541 ctctgtgtctgaagaaacaaagctatgacgtcattgtcctatcttgccaaattgtgacag 600
Oy 121 aaggaagacttgaagaaagcagaagaaaggaaggaagcactgaagcttggggccaaaaggcttc 180
    |||||||
Db 601 aaggaagacttgaagaaagcagaagaaaggaaggaagcactgaagcttggggccaaaaggcttc 660
Oy 181 attgaagatcagcaagagagttgtgagagagttcatcttgcgcggccatccagtcacgc 240
    |||||||
Db 661 attgaagatcagcaagagagttgtgagagagttcatcttgcgcggccatccagtcacgc 720
Oy 241 gcaatgatatgagacgcgtactcctctgtgagcactctcttgcgaagcctgacatgcgccgc 300
    |||||||
Db 721 gcaatgatatgagacgcgtactcctctgtgagcactctcttgcgaagcctgacatgcgccgc 780
Oy 301 aaacaagtgaaatcgcccgagcgaggaagggccaagtatgtgtccacgagcgccacagga 360
    |||||||
Db 781 aaacaagtgaaatcgcccgagcgaggaagggccaagtatgtgtccacgagcgccacagga 840
Oy 361 aagggaaacgatacagctcgtttgagtcagctgactctacactgagcccccagataaag 420
    |||||||
Db 841 aagggaaacgatacagctcgtttgagtcagctgactctacactgagcccccagataaag 900
Oy 421 gcaatgtcccttgagagatgcctgaaattctaaacagggttcaagggccgcaatgacctg 480
    |||||||
Db 901 gcaatgtcccttgagagatgcctgaaattctaaacagggttcaagggccgcaatgacctg 960
Oy 481 atgagatcagcaagcaacagagattccatcccggttcaactcccaagaacccgtggagc 540
    |||||||
Db 961 atgagatcagcaagcaacagagattccatcccggttcaactcccaagaacccgtggagc 1020

```

```

? PRIOR APPLICATION NUMBER: US 09/560,875
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: PCT/US01/03800
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 09/515,126
? PRIOR FILING DATE: 2000-02-28
? PRIOR APPLICATION NUMBER: US 09/577,409
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/04927
? PRIOR FILING DATE: 2001-02-26
? PRIOR APPLICATION NUMBER: US 09/519,705
? PRIOR FILING DATE: 2000-03-07
? PRIOR APPLICATION NUMBER: US 09/574,454
? PRIOR FILING DATE: 2000-05-19
? PRIOR APPLICATION NUMBER: PCT/US01/04941
? PRIOR FILING DATE: 2001-03-05
? PRIOR APPLICATION NUMBER: US 09/540,217
? PRIOR FILING DATE: 2000-03-31
? PRIOR APPLICATION NUMBER: US 09/649,167
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: PCT/US01/08631
? PRIOR FILING DATE: 2001-03-30
? PRIOR APPLICATION NUMBER: US 09/552,929
? PRIOR FILING DATE: 2000-04-18
? PRIOR APPLICATION NUMBER: US 09/770,160
? PRIOR FILING DATE: 2001-01-26
? PRIOR APPLICATION NUMBER: PCT/US01/08656
? PRIOR FILING DATE: 2001-04-18
? PRIOR APPLICATION NUMBER: US 09/577,408
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/14827
? PRIOR FILING DATE: 2001-05-16
? NUMBER OF SEQ ID NOS: 1192
? SOFTWARE: PL_FL-genes Version 6.0
? SEQ ID NO 330
? LENGTH: 2005
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (254)..(2005)
US-60-340-187-330

```

Query Match	Best Local	Similarity	95.5%	Score	1183.8	DB	9	Length	2005
Matches	1188	Conservative	0	Mismatches	7	Indels	0	Gaps	0
QY	1	atgtccagcaaaaggtccctcgctgtctctgtgcgtacaaagtcgagcgctggagacctgtgac	60						
Db	482	atgtccagcaaaaggtccctcgctgtctctgtgcgtacaaagtcgagcgctggagacctgtgac	541						
QY	61	ctcgtgtgtcgtgaagggaaacaaagctatgtacgttcattgcctatctgtgcacaatgtgcag	120						
Db	542	ctcgtgtgtcgtgaagggaaacaaagctatgtacgttcattgcctatctgtgcacaatgtgcag	601						
QY	121	aaggaagaacctctgcagggagacccaaggaaggaagcgtgaagccttggggcccaaaagtgcttc	180						
Db	602	aaggaagaacctctgcagggagacccaaggaaggaagcgtgaagccttggggcccaaaagtgcttc	661						
QY	181	attgagagatctgcagcgaggaattctgtgagagattcabcctctgcgcggacatccagctcacg	240						
Db	662	attgagagatctgcagcgaggaattctgtgagagattcabcctctgcgcggacatccagctcacg	721						
QY	241	gcactgtatataagaccgctaccctctcctctgggagacactctcttcgcagggccctgcagtcgccc	300						
Db	722	gcactgtatataagaccgctaccctctcctctgggagacactctcttcgcagggccctgcagtcgccc	781						
QY	301	aaacaagctggaaatcgccccagcgggaggggccaagtatgtgtcccaagcgccgacaaga	360						
Db	782	aaacaagctggaaatcgccccagcgggaggggccaagtatgtgtcccaagcgccgacaaga	841						
QY	361	aagggagagatcatagctccgtgttgtagctcaagctcgtactactactgtgccccacagataag	420						



```

Db 181 attgagatgctgcagcaggagttgtgagaggtatcatctgcccgcacatccagtcacg 240
QY 241 gacgtatagagaccgctactactccgggacac-tctcttgcagagccctgcatgcgcg 299
Db 241 gacgtatagagaccgctactactccgggacac-tctcttgcagagccctgcatgcgcg 300
QY 300 caaacagctgagaa-tcgccagcggaggggagccagatgtgtcccccgcgcacag 358
Db 301 caaacagctgagaaattcgcgcagcgaggggagccagatgtgtgtcccccgcgcacag 360
QY 359 gaaagggagaaacatcagctcgggttgagctcagctgctactactgag-cccccagata 417
Db 361 gaaagggagaaacatcagctcgggttgagctcagctgctactactgagccccccagata 420
QY 418 aaggtcattgctccctgggaggtgctgagatcttacaacccggttcaaggcgccagatgac 477
Db 421 aaggtcattgctccctgggaggtgctgagatcttacaacccggttcaaggcgccagatgac 480
QY 478 ctgattgagtaacgcaaaagc-aacacggagatcccatcccggtcactcccaagaaaccgtg 536
Db 481 ctgattgagtaacgcaaaagcattacatggaattcccatcccggtcactcccaagaaaccgtg 540
QY 537 gagcatgtgtagaaccctcacaacacacagtaagagctggaatccctgtagaaccacaa 596
Db 541 gagcatgtgtagaaccctcacaacacacagtaagagctggaatccctgtagaaccacaa 600
QY 597 gaaccaagcgctccaggtctcttacaacgaaagacagacacacacacacacacacacac 656
Db 601 gaaccaagcgctccaggtctcttacaacgaaagacacacacacacacacacacacacac 660
QY 657 cccctgacattctcgagatcgag-tcaaaaaggggtccctgtg--aaggtgacacacgt 713
Db 661 cccctgacattctcgagatcgagttcaaaaaggggtccctgtgaaaggtgagacacacgt 720
QY 714 --caagggatgacacacacacacacac-tccttggagctc-ttcatgtactgtgagag 769
Db 721 tcaaggtatgagacacacacacacacacacacacacacacacacacacacacacacac 780
QY 770 tcgcgagcaagcatgtgctgagc--gtatgacacatg-tgagaacacgcttcaat--g 823
Db 781 tcgcgagcaagcatgtgctgagcgttgagcgttgaatctgagacacacacacacacac 840
QY 824 gaattgaatgccgaggtatctc--aacgagacccacagacacacacacacacacacac 879
Db 841 aattgaatgccgaggtatctccttaccgagacacacacacacacacacacacacacac 900
QY 880 cattagacatccgagccttaccacatgagccgagagtgagcaaatc--aaacagagc 936
Db 901 cattagacatccgagccttaccacatgagccgagagtgagcaaatc--aaacagagc 960
QY 937 ctgggcttgaatctgctgagc-tggtgtataccggttacc--ggcctagaccgtgaggt 993
Db 961 ctgggcttgaatctgctgagcgtgtgataccggttcttctgacacagccctgaggt 1020
QY 994 gaattgttcgcgcacatgac--cgcaagtcaccaagagcagatgtgagaggg-aaagtga 1049
Db 1021 gaattgttcgcgcacatgac--cgcaagtcaccaagagcagatgtgagagggaaagtga 1080
QY 1050 ggtgttcgctcctc--aaaggccaggtgtgacatcc--tcggccggagagtcacacgtgc 1103
Db 1081 ggtgttcgctcctcctcaaaaggccaggtgtgatacctcctcgccggagagtcacacacgt 1140
QY 1104 tcttacaatgtagagctgtgagcaagcgtgcaaggtgattatga 1151
Db 1141 tcttctacaatgtagagctgtgagcaagcgtgcaaggtgca 1188

```

```

; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; LIBRARIES
; FILE REFERENCE: 20411-748CON1
; CURRENT APPLICATION NUMBER: US/09/898,888A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/340,623
; PRIOR FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/205,070
; PRIOR FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 22919
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(987)
; OTHER INFORMATION: n = A,T,C or G
US-09-898-888A-22919

```

```

Query Match 44.8%; Score 554.6; DB 6; Length 987;
Best Local Similarity 87.4%; Pred. No. 2,6e-142;
Matches 670; Conservative 0; Mismatches 4; Indels 93; Gaps 2;

```

```

QY 473 atgacctgattgagtagcagcaagcaaacgagattccatcccgctactcccaagaacc 532
Db 2 atgacctgattgagtagcagcaagcaaacgagattccatcccgctactcccaagaacc 61
QY 533 cgtgagcatgtagtgaagaaactctatgcacatcagatcagaggtggaatccgtggagacc 592
Db 62 cgtgagcatgtagtgaagaaactctatgcacatcagatcagaggtggaatccgtggagacc 121
QY 593 ccaagaacaaagcgctccaggtctcttacaacgaaacacaggaacacacacacacacac 652
Db 122 ccaagaacaaagcgctccaggtctcttacaacgaaacacaggaacacacacacacacac 181
QY 653 acacccctgacattctcgagatcgagttcaaaaaggggtccctgtgaggtgacacacg 712
Db 182 acacccctgacattctc----- 198
QY 713 tcaagatgtagcacaacacacacacacacacacacacacacacacacacacacacacac 772
Db 199 ----ggatgtagcacaacacacacacacacacacacacacacacacacacacacacac 254
QY 773 cgggcaagcatggtggtggtcgatgtgacatcgttggagaaacacacacacacacacac 832
Db 255 cgggcaagcatggtggtggtcgatgtgacatcgttggagaaacacacacacacacacac 314
QY 833 cccgaggtatctcagagacacacacacacacacacacacacacacacacacacacacac 892
Db 315 cccgaggtatctcagagacacacacacacacacacacacacacacacacacacacacac 374
QY 893 aggccttccatgtagcggaggtgagcaaatcaacaagaagcctgtggtgtaattgtg 952
Db 375 aggccttccatgtagcggagaa----- 397
QY 953 ctgagctgtgtataccggtttagcgctagccctgagtgtaattgttccgcacatgca 1012
Db 398 -----tgtatccggttcttgcacaaagccctgagtgtaattgttccgcacatgca 448
QY 1013 tcgcaagtcaccaagagcaggtgagagggaaagtgtcaggtgtccgtccccaaggccaag 1072
Db 449 tcgcaagtcaccaagagcaggtgagagggaaagtgtcaggtgtccgtccccaaggccaag 508
QY 1073 tgtacatcctgcgcggaggtccacacgtctcttacaatgagaggtgtgagagtgta 1132
Db 509 tgtacatcctgcgcggaggtccacacgtctcttacaatgagaggtgtgagagtgta 568
QY 1133 acgtgcaaggtgtatataagcaactgtatgcacacgggttcatcaacaatcaatccctca 1192
Db 569 acgtgcaaggtgtatataagcaactgtatgcacacgggttcatcaacaatcaatccctca 628

```

```

RESULT 6
US-09-898-888A-22919
; Sequence 22919, Application US/09898888A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.

```



QY 1193 ggctgaagaatatatctgtctccaggaagctcctcccaatag 1239  
 |||  
 Db 629 ggctgaagaatatatctgtctccaggaagctcctcccaatag 675

## RESULT 7

US-09-998-598-467/c  
 ; Sequence 467, Application US/09998598  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Chenault, Ruth A.  
 ; APPLICANT: Meagher, Madelein Joy  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121.561  
 ; CURRENT APPLICATION NUMBER: US/09/998,598  
 ; CURRENT FILING DATE: 2001-11-16  
 ; NUMBER OF SEQ ID NOS: 2606  
 ; SOFTWARE: Corixa Invention Disclosure Database  
 ; SEQ ID NO 467  
 ; LENGTH: 543  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-998-598-467

Query Match 43.7%; Score 541; DB 7; Length 543;  
 Best Local Similarity 99.6%; Pred. No. 1,2e-138;  
 Matches 541; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 393 ctgctactactgccccccagataaagctcctccctggagatgctgaattcta 452  
 |||  
 Db 543 ytgctactactgccccccagataaagctcctccctggagatgctgaattcta 484  
 QY 453 caacgggttcaaggcgccgaatgactgagatgagacgaacgaacggatcccat 512  
 |||  
 Db 483 caacgggttcaaggcgccgaatgactgagatgagacgaacgaacggatcccat 424  
 QY 513 ccgggttactcccaagaacccggtgagatgagatgagatgagatgagatgagat 572  
 |||  
 Db 423 ccgggttactcccaagaacccggtgagatgagatgagatgagatgagatgagat 364  
 QY 573 ggtcggatccctgagagaccccaagaacgaacgctccagctctctacagagacca 632  
 |||  
 Db 363 ggtcggatccctgagagaccccaagaacgaacgctccagctctctacagagacca 304  
 QY 633 ggaaccgaaccccaagaaccccaagaaccccaagaaccccaagaaccccaaga 692  
 |||  
 Db 303 ggaaccgaaccccaagaaccccaagaaccccaagaaccccaagaaccccaaga 244  
 QY 693 ccctgtgaagtgagacgaacgctcgaagatgagaccccaagaaccccaagaac 752  
 |||  
 Db 243 ccctgtgaagtgagacgaacgctcgaagatgagaccccaagaaccccaagaac 184  
 QY 753 catgtacttgaaagtcgagcggaacgagatgagcggtatgagacgctgagaa 812  
 |||  
 Db 183 catgtacttgaaagtcgagcggaacgagatgagcggtatgagacgctgagaa 124  
 QY 813 ccgcttcatggaatgagatcccgagatgagaccccaagaaccccaagaaccc 872  
 |||  
 Db 123 ccgcttcatggaatgagatcccgagatgagaccccaagaaccccaagaaccc 64  
 QY 873 catgtacttgaaagtcgagcggtatgagacgctgagaa 932  
 |||  
 Db 63 catgtacttgaaagtcgagcggtatgagacgctgagaa 4  
 QY 933 agg 935  
 |||  
 Db 3 AGG 1

## RESULT 8

US-09-998-598-702  
 ; Sequence 702, Application US/09998598  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Chenault, Ruth A.  
 ; APPLICANT: Meagher, Madelein Joy  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
 ; FILE REFERENCE: 210121.561  
 ; CURRENT APPLICATION NUMBER: US/09/998,598  
 ; CURRENT FILING DATE: 2001-11-16  
 ; NUMBER OF SEQ ID NOS: 2606  
 ; SOFTWARE: Corixa Invention Disclosure Database  
 ; SEQ ID NO 702  
 ; LENGTH: 503  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-998-598-702

Query Match 40.6%; Score 503; DB 7; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 3,2e-128;  
 Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 ctgctactactgccccccagataaagctcctccctggagatgctgaattcta 452  
 |||  
 Db 1 ctgctactactgccccccagataaagctcctccctggagatgctgaattcta 60  
 QY 453 caacgggttcaaggcgccgaatgactgagatgagacgaacgaacggatcccat 512  
 |||  
 Db 61 caacgggttcaaggcgccgaatgactgagatgagacgaacgaacggatcccat 120  
 QY 513 ccgggttactcccaagaacccggtgagatgagatgagatgagatgagatgagat 572  
 |||  
 Db 121 ccgggttactcccaagaacccggtgagatgagatgagatgagatgagatgagat 180  
 QY 573 ggtcggatccctgagagaccccaagaacgaacgctccagctctctacagagacca 632  
 |||  
 Db 181 ggtcggatccctgagagaccccaagaacgaacgctccagctctctacagagacca 240  
 QY 633 ggaaccgaaccccaagaaccccaagaaccccaagaaccccaagaaccccaaga 692  
 |||  
 Db 241 ggaaccgaaccccaagaaccccaagaaccccaagaaccccaagaaccccaaga 300  
 QY 693 ccctgtgaagtgagacgaacgctcgaagatgagaccccaagaaccccaagaac 752  
 |||  
 Db 301 ccctgtgaagtgagacgaacgctcgaagatgagaccccaagaaccccaagaac 360  
 QY 753 catgtacttgaaagtcgagcggtatgagacgctgagaa 812  
 |||  
 Db 361 catgtacttgaaagtcgagcggtatgagacgctgagaa 420  
 QY 813 ccgcttcatggaatgagatcccgagatgagaccccaagaaccccaagaaccc 872  
 |||  
 Db 421 ccgcttcatggaatgagatcccgagatgagaccccaagaaccccaagaaccc 480  
 QY 873 catgtacttgaaagtcgagcggtatgagacgctgagaa 895  
 |||  
 Db 481 catgtacttgaaagtcgagcggtatgagacgctgagaa 503

## RESULT 9

US-09-998-598-907  
 ; Sequence 907, Application US/09998598  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Chenault, Ruth A.  
 ; APPLICANT: Meagher, Madelein Joy  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER

```

; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 907
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-907

```

```

Query Match          40.6%; Score 503; DB 7; Length 503;
Best Local Similarity 100.0%; Pred. No. 3.2e-128;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 393 ctgctactactgccccccagataaagtcattctccctggagatgctgaattcta 452
DB 1 ctgctactactgccccccagataaagtcattctccctggagatgctgaattcta 60
QY 453 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgagatcccat 512
DB 61 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgagatcccat 120
QY 513 ccggttactcccaagaaccgctggagcatgagtagaacctctacatcagctacga 572
DB 121 ccggttactcccaagaaccgctggagcatgagtagaacctctacatcagctacga 180
QY 573 ggcctgaatcctctggaagaccacgaagacgctctcgaagtctctacagaagacca 632
DB 181 ggcctgaatcctctggaagaccacgaagacgctctcgaagtctctacagaagacca 240
QY 632 ggaaccagccaaagcccccaacacccctgacatctcgaagtcgaagttcaaaaagggt 692
DB 241 ggaaccagccaaagcccccaacacccctgacatctcgaagtcgaagttcaaaaagggt 300
QY 693 cccctgtaaggtgacaaagtcgaagtagcgaacacccacgacccctcttggagctct 752
DB 301 cccctgtaaggtgacaaagtcgaagtagcgaacacccacgacccctcttggagctct 360
QY 753 catgtacctgaagcgaagtcgagcgaagcattgagctgagctcttgaatcgtagagaa 812
DB 361 catgtacctgaagcgaagtcgagcgaagcattgagctgagctcttgaatcgtagagaa 420
QY 813 ccgcttcatggaatggaagtcgagcgaagtagctacacgaaccccgagcagaccatctta 872
DB 421 ccgcttcatggaatggaagtcgagcgaagtagctacacgaaccccgagcagaccatctta 480
QY 873 ccagtcatttagacatcgag 895
DB 481 ccagtcatttagacatcgag 503

```

## RESULT 10

```

; Sequence 924, Application US/09998598
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 924
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-924

```

```

Query Match          40.6%; Score 503; DB 7; Length 503;
Best Local Similarity 100.0%; Pred. No. 3.2e-128;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 393 ctgctactactgccccccagataaagtcattctccctggagatgctgaattcta 452
DB 1 ctgctactactgccccccagataaagtcattctccctggagatgctgaattcta 60
QY 453 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgagatcccat 512
DB 61 caaccggttcaaggccgcaatgacatgagtagcgaagcaacacgagatcccat 120
QY 513 ccggttactcccaagaaccgctggagcatgagtagaacctctacatcagctacga 572
DB 121 ccggttactcccaagaaccgctggagcatgagtagaacctctacatcagctacga 180
QY 573 ggcctgaatcctctggaagaccacgaagacgctctcgaagtctctacagaagacca 632
DB 181 ggcctgaatcctctggaagaccacgaagacgctctcgaagtctctacagaagacca 240
QY 632 ggaaccagccaaagcccccaacacccctgacatctcgaagtcgaagttcaaaaagggt 692
DB 241 ggaaccagccaaagcccccaacacccctgacatctcgaagtcgaagttcaaaaagggt 300
QY 693 cccctgtaaggtgacaaagtcgaagtagcgaacacccacgacccctcttggagctct 752
DB 301 cccctgtaaggtgacaaagtcgaagtagcgaacacccacgacccctcttggagctct 360
QY 753 catgtacctgaagcgaagtcgagcgaagcattgagctgagctcttgaatcgtagagaa 812
DB 361 catgtacctgaagcgaagtcgagcgaagcattgagctgagctcttgaatcgtagagaa 420
QY 813 ccgcttcatggaatggaagtcgagcgaagtagctacacgaaccccgagcagaccatctta 872
DB 421 ccgcttcatggaatggaagtcgagcgaagtagctacacgaaccccgagcagaccatctta 480
QY 873 ccagtcatttagacatcgag 895
DB 481 ccagtcatttagacatcgag 503

```

## RESULT 11

```

; Sequence 1365, Application US/09998598
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1365
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1365

```

```

Query Match          40.3%; Score 499.8; DB 7; Length 503;
Best Local Similarity 99.6%; Pred. No. 2.4e-127;
Matches 501; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 393 ctgctactactgccccccagataaagtcattctccctggagatgctgaattcta 452
DB 503 CTGCTACTACTGCCCCCAGATAAAGTCATTCTCTCGAGATGCTGAATTCCTA 444

```

Db	181	ctctccaggtctctacagagaccacccaggaaccagccaagggcccccacaccccttgacatt	240
Qy	667	ctcgagatcgaftctcaaaaaggggtcccttgaaagtgaccaagttaaagtatggcacc	726
Db	241	ctcgagatcgagftctcaaaaaggggtcccttgaaagtgaccaagttcaagaaatggcacc	300
Qy	727	accacacagacgtccctctgagctctctacgtacacgtacgaagtcgcggccaagcatggc	786
Db	301	accacacagacgtccctctgagctctctacgtacacgtacgaagtcgcggccaagcatggc	360
Qy	787	gtgggctcgatctaacatctgtgagaagacgcgtcttcttgaaatgagtcgccgagttctac	846
Db	361	gtgggctcgatctaacatctgtgagaagacgcgtcttcttgaaatgagtcgccgagttctac	420
Qy	847	gagaccacagcgacacactcctttacacatgctcatttagacatcgagg	895
Db	421	gagaccacagcgacacactcctttacacatgctcatttagacatcgagg	469
RESULT 13			
US-09-933-524A-54674			
: Sequence 54674, Application US/09933524A			
: GENERAL INFORMATION:			
: APPLICANT: Drmanac, Radje T.			
: APPLICANT: Labat, Ivan			
: APPLICANT: Stacks-Crain, Birgit			
: APPLICANT: Dickson, Mark			
: APPLICANT: Jones, Lee W.			
: TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained			
: TITLE OF INVENTION: From Various Libraries			
: FILE REFERENCE: 774			
: CURRENT APPLICATION NUMBER: US/09/933,524A			
: FILING DATE: 2001-08-20			

```

RESULT 13
US-09-933-524A-54674
? Sequence 54674, Application US/0933524A
? GENERAL INFORMATION:
? APPLICANT: Dimanac, Radoje T.
? APPLICANT: Labat, Ivan
? APPLICANT: Stache-Crain, Birgit
? APPLICANT: Dickson, Mark
? APPLICANT: Jones, Lee W.
? TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
? TITLE OF INVENTION: From Various Libraries
? FILE REFERENCE: 774
? CURRENT APPLICATION NUMBER: US/09/933, 524A
? CURRENT FILING DATE: 2001-08-20
? PRIOR APPLICATION NUMBER: 09/528, 409
? PRIOR FILING DATE: 2000-05-17
? NUMBER OF SEQ. ID NOS: 116231
? SOFTWARE: HY-patent.pl Version 3.1
? SEQ ID NO 54674
? LENGTH: 479
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-933-524A-54674

Query Match 36.9%; Score 457.4; DB 6; Length 479;
Best Local Similarity 99.8%; Pred. No. 1,1e-115;
Matches 458; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 3 gtccagcaaaagctccgtggtctctgtgacctacagtggcgccttggaacccctgfgaatct 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20 gtccagcaaaagctccgtggtctctgtgacctacagtggcgccttggaacccctgfgaatct 79
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 gctgtgtgcttgaaggaacaagagctctgacgtcatctgtccatctgtgccaacattggccagaa 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Query Match	36.9%	Score 457.4	DB 6	Length 4/9
Best Local Similarity	99.8%	Pred. No. 1,1e-115		
Matches 458	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	3	gtccagcaaaagctccggtgtctctgacctaaagtgcgcgcttgacaacctgtgacacct	62	
Db	20	gtccagcaaaagctccggtgtctctgacctaaagtgcgcgcttgacaacctgtgacacct	79	
QY	63	cgtgtgctgaaggagacaagactatgacgtcaattgtctactctgccaactttggccaagaa	122	
Db	80	cgtgtgtgctgaaggagacaagactatgacgtctgtctactctgccaactttggccaagaa	139	
QY	123	ggaaagactctgcagaggaagccaaggaagagcactgaagtttggggccaaaaggtgttcat	182	
Db	140	ggaaagactctgcagaggaagccaaggaagagcactgaagtttggggccaaaaggtgttcat	199	
QY	183	ttagagcttcagcagcagggattgttggaagagttaactctgcgcgcgcacatccagtcacagc	242	
Db	200	ttagagcttcagcagcagggattgttggaagagttaactctgcgcgcgcacatccagtcacagc	259	
QY	243	actggtatggagacggctactcctctgcgcacccctctcttgcagggcccttgcacatcgcccgcaa	302	
Db	260	actggtatggagacggctactcctctgcgcacccctcttgcagggcccttgcacatcgcccgcaa	319	
QY	303	acaagaatggaaatcgcccaagcggagaggggccaagatgtgttccacagcgccacacagaa	362	
Db	320	acaagaatggaaatcgcccaagcggagaggggccaagatgtgttccacagcgccacacagaa	379	
QY	363	ggggaaagatcatgctgcggtttgagctcagctgtctactcactcggccccacagataaagtg	422	

Db 380 ggggaacgacatcagctcgtttgagctcagctcgtctactactcagcccccagaataagtt 439  
 Qy 423 catgtctcctgagagatccctgatatctcaaccggtt 461  
 Db 440 catgtctcctgagagatcgtctgaaattctcaaccggtt 478

## RESULT 14

US-10-029-386-8869/c  
 ; Sequence 8869, Application US/10029386  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
 ; FILE REFERENCE: AEMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; NUMBER OF SEQ ID NOS: 3428  
 ; SOFTWARE: Annomax Sequence Listing Engine Vers. 1.1  
 ; SEQ ID NO 8869  
 ; LENGTH: 538  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC005000.1  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.76  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
 ; OTHER INFORMATION: NE HIT: g114740527, EVALU0.00e+00  
 ; OTHER INFORMATION: SWISSPROT HIT: P00966, EVALU0.00e+00  
 ; OTHER INFORMATION: EST\_HUMAN HIT: BE783168.1, EVALU0.00e+00  
 US-10-029-386-8869

## Query Match

Best Local Similarity 36.6%; Score 454; DB 8; Length 538;  
 Matches 492; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Qy 104 tggccacatttggccagaagaagacttggagaagccaggaagaagccactgaagcttg 163  
 Db 538 TGCCCAACACTGGCCAGAGGAAGACTTTGAGGAAGCCAGAGAGAGCCACTGAGCTGG 479  
 Qy 164 gggccaaaaggtgttcattgagatgttcagcagggagttgttgaggagttcattggc 223  
 Db 478 GGGGCAAAAAGTGTTCATTGAGAGAGTCAGCAAGAGACTTGTGAAGAGTTTCATCTGGC 419  
 Qy 224 cggccatcagtcagcgacatgtagagacgctcctctctgggacccctcttgca 283  
 Db 418 TGCCCATTCAGTCACAGCGCACTGTATGAGAGCACTACTCTCTGCACTCAACA 359  
 Qy 284 gggccctgcatcggccgcaaaaagtggaatcccgccagcgggaggggccaagtatgt 343  
 Db 358 GGCCCTGCATCCCGCAAAACAGAGGAATCGCCAGCAGAGAGGAGGCAAGTAAGTGT 299  
 Qy 344 cccacggcgccacgaaggaagggaacagatcagtcggtttgagactaagttgttaccac 403  
 Db 298 CCCACAGCGCCACAGAGGAAGGAATGATCAGTTCAGTTGAACTTAACCTGCTACTCTC 239  
 Qy 404 tgg-ccccccagataaagttcattgtcctccttgagagatgctgaattctcaaccggttc 462  
 Db 238 TGGCCCCCCCCAGATTAAGTATGCTCCCTGGAGATGCCCAAGTTCTACACAGGTTTC 179  
 Qy 463 aaggcgccaaatgacctgtagtaacgaaagcaaacagagagttccatcccggttcaact 522  
 Db 178 AAGGTCCGAATGACCTTATGGAACACACAAACGACACGAGATTCCTCCATCCGAGTCACT 119  
 Qy 523 cccaaagaccggtgagcagatgtagtaacatcagacatcagcagcagcagcagcagcagc 582  
 Db 118 CCCAAGAACCTGTGAGACATGAGACGAGAACCTCATGACATGCAATAGAGCTGGAATC 59

Qy 583 ctggagaacccccagaaccagcgcctcaggttctacacgaagaccagagaccag 640  
 Db 58 TTGGAGAACCTTGAAGAACCAAGCATTTCCAGTCTCTCAGAGAACCCAGAGACCGG 1

## RESULT 15

US-09-922-340-266  
 ; Sequence 266, Application US/09922340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
 ; FILE REFERENCE: 20411-726CON2  
 ; CURRENT APPLICATION NUMBER: US/09/922,340  
 ; PRIOR APPLICATION NUMBER: 09/353,690  
 ; PRIOR FILING DATE: 1999-07-14  
 ; PRIOR APPLICATION NUMBER: US 09/004,182  
 ; PRIOR FILING DATE: 1998-01-07  
 ; NUMBER OF SEQ ID NOS: 12181  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 266  
 ; LENGTH: 465  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1)...(465)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-922-340-266

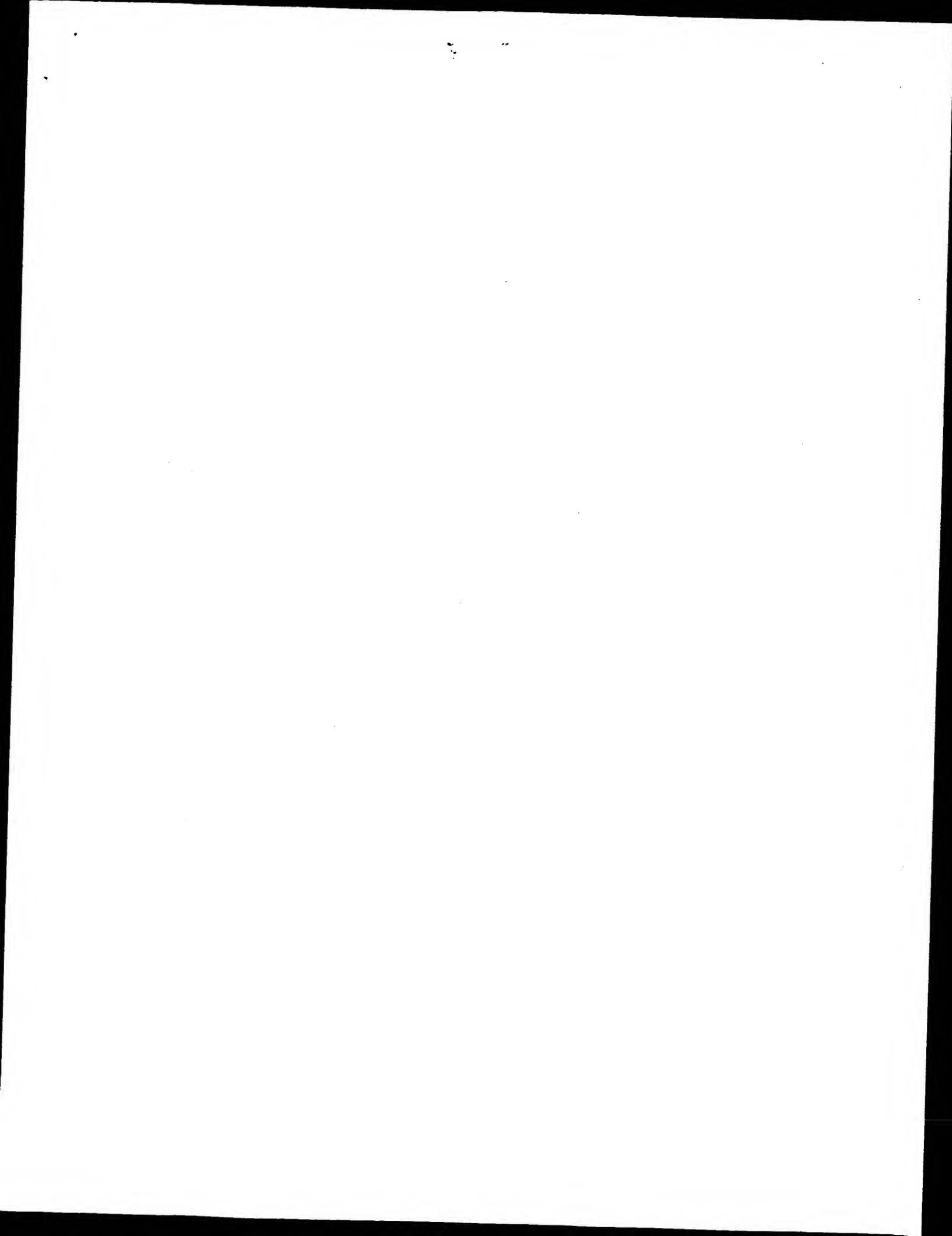
## Query Match

Best Local Similarity 36.6%; Score 453.2; DB 6; Length 465;  
 Matches 455; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 334 aagtaatgttccacagcgccacaggaagaagggaagatcagctcgtttgagctcagc 393  
 Db 1 aagtaatgttccacagcgccacaggaagaagggaagatcagctcgtttgagctcagc 60  
 Qy 394 tgcatactctggccccagataaagttcattgtctccttgagagatgagctgaattctac 453  
 Db 61 tgcatactctggccccagataaagttcattgtctccttgagagatgagctgaattctac 120  
 Qy 454 aaccggttcaagggccgcaatgacatgtagtagtagtagtagtagtagtagtagtagtag 513  
 Db 121 aaccggttcaagggccgcaatgacatgtagtagtagtagtagtagtagtagtagtagtag 180  
 Qy 514 ccggtatcctcccaagaacccgttgtagtagtagtagtagtagtagtagtagtagtagtagtag 573  
 Db 181 ccggtatcctcccaagaacccgttgtagtagtagtagtagtagtagtagtagtagtagtagtag 240  
 Qy 574 gcttgaatcctggagaaaccccaagaacccgctccaggtctctacacgaagaccag 633  
 Db 241 gcttgaatcctggagaaaccccaagaacccgctccaggtctctacacgaagaccag 300  
 Qy 634 gaccagcgaagcccccacaacacccctgacatctcagagatcaggttcaaaaagggttc 693  
 Db 301 gaccagcgaagcccccacaacacccctgacatctcagagatcaggttcaaaaagggttc 360  
 Qy 694 ccttgaagttgacaaagctcaagatgtagtagtagtagtagtagtagtagtagtagtagtagtag 753  
 Db 361 ccttgaagttgacaaagctcaagatgtagtagtagtagtagtagtagtagtagtagtagtagtag 420  
 Qy 754 atgtacttgaaggaagtcgaggggcaagcagcagcagcagcagcagcagcagcagcagcagc 792  
 Db 421 atgtacttgaaggaagtcgaggggcaagcagcagcagcagcagcagcagcagcagcagcagc 459

Wed Feb 13 07:36:05 2002

us-09-775-693-1.rnpn





```

117 yAlaThrGlyGlyAsnAspValArgPheGluLeuSerCysTyr 134
401 CACTGGCCCCCGAATAAGGTCAATGCTCCCTGGAGATGCTCAATTC 450
134 erLeuAlaProGlnIleLysValIleAlaProTyrArgMetProGluPhe 150
451 TACACCGGTTCAAGGGCCCAATGACCTGATGGAGTCCGAAGACACA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnI 167
501 CGGATTCCTCATCCCGGTACTCCCAAGAACCCGTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysAsnProTyrSerMetAspGlu 184
551 ACCATGACATGACATGACGATGAGGCTGATTCGAGAACCCCAAGAAC 600
184 snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAAGCGCTCCAGGTCTCTACACGAGAACCCAGAGCCCAAGCCCAAGCCC 650
201 GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATCTCCAGATCGAGTCAAAAAAGGGGCTCTGTA 700
217 oAsnThrProAspIleLeuGlnIleGluPheLysGlyValProValI 234
701 AGGTGACCAACGTCAAGATGACACCCAGACACCTCCCTGGAGCTC 750
234 yValAlThrAsnValLysAspGlyThrThrHisIleThrSerLeuGluLeu 250
751 TTCATGACCTGAACGAACTGCGGGCCAAACATGCGGTGGCCGATTTGA 800
251 PheMetTyrLeuAsnGlnValAlaGlyLysHisGlyValGlyArgIleAs 267
801 CATCTGTGAGAACCGCTTCATTTGAAATGAAGTCCGAGTATCTACAGA 850
267 rIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGluT 284
851 CCCACAGGACCAATCCTTTACCATGCTCATTTAGACATCGAGCCCTTC 900
284 hProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 300
901 ACCATGACCGGGAAGTGGCAAAATCAACAAAGGCTGGGCTTGAATTT 950
301 ThrMetAspArgGluValArgLysIleLysGlnGlyLeuLysPhe 317
951 TCGTACGCTGGTGTATACCGGTTACGGCTTACGCTTGAATTTG 1000
317 eAlaGluLeuValTyrThrGlyLeuArgProSerProGluCysGluPhe 334
1001 TCCGCGCATGCTCGCAAGTCCAGGAGGAGGAGGAAAGGAAAGTCCAG 1050
334 aArgHisCysIleIleAlaLysSerGlnGlnArgValGluGlyLysValGln 350
1051 GTTCCGCTCCCAAGGCGCAGGTATACATCTCGCGCGGAGTCCCACT 1100
351 ValSerValLeuLysGlyGlnValTyrIleLeuGlyArgGluSerProLe 367
1101 GTCTCTACATGAGAGCTGTGAGCATGAACGTGACAGGTGATTTAG 1150
367 uSerLeuTyrAsnGlnGluLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCACTGATGACACCGGTTTCATCAACATCAATTCCTCAGCGGTGAG 1200
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTCCAGAGCATGCTCACTGCCAAA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-775-693-7

```

```

seq documentation block:
<sequence 7> Application 05/09/75693-1
GENERAL INFORMATION:
APPLICANT: Clark, Mike
APPLICANT: Holtzberg, Frederick Wayne
APPLICANT: Ensor, Charles Mark
TITLE OF INVENTION: Methods for Predicting Susceptibility of Patients to Arginine
FILE REFERENCE: PHOS-0060
CURRENT APPLICATION NUMBER: US/09/775,693
CURRENT FILING DATE: 2001-02-02
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 7
LENGTH: 412
TYPE: PRT
ORGANISM: Homo sapiens
US-09-775-693-7

alignment_scores:
Quality: 2144.00 Length: 412
Ratio: 5.204 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-693-1 x US-09-775-693-7 ..
Align seq 1/1 to: US-09-775-693-7 from: 1 to: 412

1 ATGTCCAGCAAAAGGCTCCGTGGTTCGTGCGCTACAGTGGCGGCTGACAC 50
1 MetSerSerLysGlySerValValLeuAlaTyrSerGlyGlyLeuAspThr 17
51 CTGTGCATCCCTGCTGGCTGAGACAGACAGGCTATACGTCATTTGCT 100
17 rSerCysIleLeuValTyrPheLysGlnGluIleGlnIleTyrAspValIleAla 34
101 ATCTGGCCACATTTGGCCAGAAAGACTTCGAGAGAGCCAGAGCAAGAG 150
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGluAlaArgLysLys 50
151 GCACGAACTGGGGCCCAAAAGGTGTTCAATTGAGATGTCACAGAGA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnLysPheValSerArgI 67
201 GTTTGTGAGAGATTCATCTGGCGGCGCATCCAGTCCAGGCACTGTATG 250
67 uPheValGlnGluPheIleTyrProAlaIleGlnSerSerAlaLeuTyrG 84
251 AGGACCGTACCTCTCTGGGCACTCTCTTGGCAGGCCCTGCATCGCCCG 300
84 LuAspArgTyrLeuLeuGlnThrSerLeuAlaArgProCysIleAlaArg 100
301 AAACAGTGAATCGCCCAAGCGGAGGAGGCGCAAGATGTGTGCCACGG 350
101 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 117
351 CGCAGCAGAAAGGAGACATCAGGTCCGCTTGTGAGTCACTGCTGCTACT 400
117 yAlaThrGlyLysGlyAsnAspValValArgPheGluLeuSerCysTyrS 134
401 CACTGGCCCCCGAATAAGGTCAATGCTCCCTGGAGATGCTCAATTC 450
134 erLeuAlaProGlnIleLysValIleAlaProTyrArgMetProGluPhe 150
451 TACACCGGTTCAAGGGCCCAATGACCTGATGGAGTCCGAAGACACA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnI 167
501 CGGATTCCTCATCCCGGTACTCCCAAGAACCCGTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysAsnProTyrSerMetAspGlu 184

```



```

551 ACCATGACATGACAGCTACGAGGCTGGAGATCTTGAGAACCCCAAGAAC 600
|||||
184 snleuethisileserlyrgluialaglylleugluasnprolyasn 200
601 CAAGGGCCCGAGGTCTCTACAGAGAACCCCAAGGCCCAAGGCC 650
|||||
201 GluAlaProProGlyLeuThrlystrGlnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATCTCTGAGATGAGTTCAAAAAGGGCTCCCTGGA 700
|||||
217 OAsnThrProAspIleLeuGluIleGluPheLysGlyValAlProVal 234
701 AGGTGACCAAGCTGCAAGATGGCCACCACCACCAAGACCTCTTGAGCTC 750
|||||
234 yValAlThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGluLeu 250
751 TTCATGTACTGACAGAACTCCGGGCAAGCATGGCGTGGCGGTATTGA 800
|||||
251 PheMetTyLLeuAsnGluValAlaGlyLysHisGlyValAlaArgIleAs 267
801 CATCTGGAGAACCGCTTCATGTGAATGAAGTCCGAGGTATCTACGAGA 850
|||||
267 rIleValAlGluAsnArgPheIleGlyMetLysSerArgGlyIleTyrgLut 284
851 CCCCAAGCAGGACCATCTTACATGCTCATTTAGACATCGAGGCCCTTC 900
|||||
284 hProAlaGlyThrIleLeuTyHisAlaHisLeuAspIleGluAlaPhe 300
901 ACCATGACAGGGAAGTGGGCAAAATCAACAGAGCTGGCTTGAATTT 950
|||||
301 ThrMetAspArgGluValAlaArgLysIleLysGlnGlyLeuGlyLeuLysSh 317
951 TCCTGAGCTGGTGTATACCGGTTTACGGCTAGCCCTGAGTGTGAATTTG 1000
|||||
317 eAlaGluLeuValTyThrGlyLeuArgProSerProGluLysGluPheY 334
1001 TCCGCCATGTCATCGCCAAAGTCCCAAGAGCGAGTGAAGGGAAGTGCAG 1050
|||||
334 aAlaArgIscysIleAlaLysSerGlnGluArgValGlnGlyLysValGln 350
1051 GTGTCCGTCTCAAGGGCCAGGCTATCATCTCTGCGCGGAGTCCCACT 1100
|||||
351 ValSerValLeuLysGlyGlnValTyThrLeuGlyAlaGlyLysProbe 367
1101 GTCCTCTCAATGAGAGCTGTGAGCATGAACGTGACAGGGGTATTAG 1150
|||||
367 uSerLeuTyAsnGluGluLeuValSerMetLysValGlnGlyAspTyrg 384
1151 AGCAAGTATGCAACCGGTTTCATCAATCAATCAATCCCTCAGGCTGAAG 1200
|||||
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATGCTCTCCAGAGCAAGTCACTGCCAAA 1236
|||||
401 GluTyHisArgLeuGlnSerLysValThrAlaLys 412
seq_name: /cgn2_6/ptodata/2/paa/us60_COMB.pep:us-60-239-841-87
seq_documentation_block:
: Sequence 87, Application US/60239841
: GENERAL INFORMATION:
: APPLICANT: Lasek, Amy W.
: APPLICANT: Jones, David A.
: TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
: FILE REFERENCE: PA-0038 P
: CURRENT APPLICATION NUMBER: US/60/239, 841
: NUMBER OF SEQ ID NOS: 194
: SOFTWARE: PERL Program
: SEQ ID NO 87
: LENGTH: 412
: TYPE: PRT

```

```

: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 1543330CD1
US-60-239-841-87

```

```

alignment_scores:
  Quality: 2123.00      Length: 412
  Ratio: 5.178          Gaps: 0
  Percent Similarity: 99.515  Percent Identity: 99.272

```

alignment\_block:

US-09-775-693-1 x US-60-239-841-87 ..

Align seg 1/1 to: US-60-239-841-87 From: 1 to: 412

```

1 ATGTCCAGCAAGGCTCCGTGCTCTGGCTACAGTGGCGGCTGGACAC 50
|||||
1 MetSerSerLysGlySerValValLeuAlaTySerIlyGlyLeuAspTh 17
51 CTGTCATCTCTCGTGTGCTGAAGAAAGGCTATGAGTCACTTGTCT 100
|||||
17 rSerCysIleLeuValItrPheLysGlnGlnGlyTyTrAspValIleAla 34
101 ATCTGGCCACATTTGGCCAGAGAGACTTTCAGAGAACCCAGAGAGAAG 150
|||||
34 yLeuAlaAsnIleGlyGlnLysGluAspPheGlnGluAlaArgLysLys 50
151 GCATGGAAGCTTGGGGCCAAAAGGTCTCATTTGAGATGTCCACAGAGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerArgG 67
201 GTTGTGAGAGAGTTCATCTGCGCGGCCCATTCAGTCCAGCGCATGTATG 250
|||||
67 uPheValGluLeuPheIleItrProAlaIleGlnSerSerAlaLeuTyrg 84
251 AGGACCGCTACCTCTGGGCACTCTCTTCGCAAGGCCCTGCATCGCCGC 300
|||||
84 LuAspArgTyLeuLeuGlnGlyThrSerLeuAlaArgProCysIleAlaArg 100
301 AAACAAGTGAATTCGCCACAGGGAGGGGCCAAGTATGTGTCCACAGG 350
|||||
101 LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyValSerHisG 117
351 CGCCACAGGAAGGGGAACGATCAGTCCGCTTGTGAGTCACTGCTACT 400
|||||
117 yAlaThrIlyLysGlyAsnAspGlnValArgPheGluLeuSerCysTyrs 134
401 CACTGGCCCCCAGATAAAGTCAATTCCTCCCTGAGAGATGCCGTAATTC 450
|||||
134 erLeuAlaProGlnIleLysValIleAlaProItrPargMetProGluPhe 150
451 TACAAACGGTTCAGAGGCCGCAATGACCTGATGAGTACGCAAAAGCAACA 500
|||||
151 TyrsAsnArgPheLysGlyArgAsnAspLeuMetGluTyThrAlaLysGln 167
501 CGGATTCCTCCATCCCGGTCACTCCAGAACCCGTGAGCATGGATGAGA 550
|||||
167 sGlyIleProIleProValItrProLysAsnProItrSerMetAspIua 184
551 ACCTCATGACATGACGATGAGGCTGGAATCTCGAGAAACCCCAAGAAC 600
|||||
184 snleuethisileserlyrgluialaglylleugluasnprolyasn 200
601 CAAGGGCTCCAGGTCTCTACAGAGAACCCAGAGCCCAAGGCC 650
|||||
201 GluAlaProProGlyLeuTyThrLysThrGlnAspProAlaLysAlaPr 217
651 CAACACCCCTGACATCTCTGAGATGAGTTCAAAAAGGGCTCCCTGGA 700
|||||
217 OAsnThrProAspIleLeuGluIleGluPheLysGlyValAlProVal 234

```

```

701 AGGTGACCAACGTCAGAGATGGACACACCCAGACACTCTTGAGCTC 750
      |||||||
234 ysvaIthrasnvallysaspGlyThrThrIstInthrSerLeuGluLeu 250
      |||||||
751 TTATGATTAAGTGAACGAGTCGCGGAGACGATGCGTGGCCGATTTGA 800
      |||||||
251 PheMetCylLeuasnGluValAlaGlyLysIstGlyValAlaGlyIleAs 267
      |||||||
801 CATCGTGGAGAACGCTTTCATTTGAATGAAATCCCGAGTATCTAGAGAG 850
      |||||||
267 pIleValGluAsnThrPheIleuLeuMetCylSerArgGlyIleuLeu 284
      |||||||
851 CCCGACGAGGACACCATCTTTACCATGCTCATTTAAGACATGAGGCTTC 900
      |||||||
284 hrProAlaGlyThrIleLeuThrIstAlaIstIleuAspIleGluAlaPhe 300
      |||||||
901 ACCATGAGACGGGAGAGTGGCCAAATCAACAAGGCTGGGCTTTGAAT 950
      |||||||
301 ThrMetAspArgGluValAlaArgLysIleLysGlnGlyLeuGlyLeuLys 317
      |||||||
951 TGGTGAAGCTGTGATACGGTTTACGGCTAGCCCTGAGTGTGAAATTG 1000
      |||||||
317 eAlaGluLeuValIleThrGlyPheThrPheIstSerProGluCysGluPhe 334
      |||||||
1001 TCCGCACTGCATCGCCAAAGTCCCGAGGACGAGTGAAGGAAAGTGCAG 1050
      |||||||
334 alArghIstCysIleAlaLysSerGlnGluValArgGlyLysValGln 350
      |||||||
1051 GTGTCCGCTCTCAAGGGCCAGGTGTACATCTCGCCCGGAGTCCCACT 1100
      |||||||
351 ValSerAlaLeuLysGlyGlnValIleLeuGlyArgGluSerProLe 367
      |||||||
1101 GTCTCTTACATGAGAGAGCTGTGATGAGTGAAGTGAAGGATGATTA 1150
      |||||||
367 uSerLeuThrAsnGluLeuValSerMetAsnValGlnGlyAspTyrG 384
      |||||||
1151 AGCCAACTGATGCCACCGGTTTCATCAACATCAATTCCTCAGCTGAAG 1200
      |||||||
384 IupProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
      |||||||
1201 GAATATCATGTCTCCAGAGCAGGTACATCGCCAAA 1236
      |||||||
401 GlutThrIstArgLeuGlnSerLysValThrAlaLys 412
seq_name: /cgn2_6/ptodata/2/paa/PCTus_COMB.pep:PCT-US00-05988-1594
seq_documentation_block:
; Sequence 1594, Application PC/TUS0005988
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Human Prostate Cancer Associated Gene Sequences and Polypeptides
; FILE REFERENCE: PA01PCT
; CURRENT APPLICATION NUMBER: PCT/US00/05988
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1594
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US00-05988-1594

```

```

alignment_scores:
  Quality: 2123.00      Length: 412
  Ratio: 5.178          Caps: 0

```

Percent Similarity: 99.515 Percent Identity: 99.272

alignment\_block:

US-09-775-693-1 x PCT-US00-05988-1594 ..

Align seg 1/1 to: PCT-US00-05988-1594 from: 1 to: 442

```

1 ATGTCCAGCAAAAGCCTCGGTGTCTGGCTCAAGTGGCGGCTGGACAC 50
      |||||||
31 MetSerSerLysGlySerValAlaLeuAlaTyrSerGlyGlyLeuAspTh 47
      |||||||
51 CTGTCGATCTCTGTGTGGCTGAAGAACAAAGGCTATGATGCTTGGCT 100
      |||||||
47 rSerCylIleuValIleuLysGlnGlyLysrAspAlaIleAlaIle 64
      |||||||
101 ATCTGGCCAACTTTGGCCAGAAAGACCTTCGAGAAAGCCAGAGAAC 150
      |||||||
64 yrLeuAlaAsnIleGlyLysGlnAspPheGlnGluAlaArgLysLys 80
      |||||||
151 GCACGTGAAGCTTGGGCGCAAAAGGTGTTCATTTGAGATGTTCAGAGGA 200
      |||||||
81 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnAspValSerArgG 97
      |||||||
201 GTTTGTGAGGAGTTCATCTGGCGCCATCCAGTCCAGCCGACCTGATG 250
      |||||||
97 uPheValGlnGluPheIleThrProAlaIleGlnSerSerIleuLys 114
      |||||||
251 AGGACCGCTACTCTCTGGGACACTCTCTTCCAGGCTCGATGCGCCG 300
      |||||||
114 IuAspArgTyrIleuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 130
      |||||||
301 AAACAAGTGAAGATGCCACGCGGAGGAGGCGCAAGTATGTATGCCACG 350
      |||||||
131 LysGlnValGluIleAlaGlnArgGluValAlaLysTyrValSerHisG 147
      |||||||
351 CGCCACAGAAAGGGGAGACGATCAAGTCCGTTTGAAGTCAAGTCTACT 400
      |||||||
147 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuSerCysTyrS 164
      |||||||
401 CACTGGCCCCCAGATTAAGTCAATGTCTCTCGAGAGATCCCTGAATTC 450
      |||||||
164 erLeuAlaProGlnIleLysValIleAlaProTyrPargMetProGluPhe 180
      |||||||
451 TACAACCGGTTCAGAGGCGGCATGACCTGATGAGTGAAGCAAGCAACA 500
      |||||||
181 TyrAsnArgPheLysGlyArgAsnAspLeuMetIuTyrAlaLysGlnHi 197
      |||||||
501 CGGATTCACATCCCGGTCACTCCCAAGAACCCGTGAGCATGATGAGA 550
      |||||||
197 sGlyIleProIleProValThrProLysAsnProTyrSerMetAspGluA 214
      |||||||
551 ACCTTCATGACATACGTCACGAGGCTGGAATCTCTGGAGAACCCCAAG 600
      |||||||
214 snLeuMetHisIleSerTyrGluAlaGlyIleLeuGluAsnProLysAsn 230
      |||||||
601 CAAGGCGCTCCAGGCTCTTACACGAGACCCAGAGCCAGCAAGGCCCC 650
      |||||||
231 GlnAlaProProGlyLeuThrIleLysThrGlnAspProAlaLysAlaPr 247
      |||||||
651 CAACACCCCTGACATTCGAGATCGAGATTCAAAAAAGGGTCCCTGTGA 700
      |||||||
247 oAsnThrProAspIleLeuGluIleGluPheLysGlyValAlaProVal 264
      |||||||
701 AGGTGACCAACGTCAGAGATGGACACACCCAGACACTCTTGAGCTC 750
      |||||||
264 ysvaIthrasnvallysaspGlyThrThrIstGlnThrSerLeuGluLeu 280
      |||||||
751 TTATGATTAAGTGAACGAGTCGCGGAGACGATGCGTGGCCGATTTGA 800
      |||||||
281 PheMetCylLeuasnGluValAlaGlyLysIstGlyValGlyArgIleAs 297
      |||||||
801 CATCGTGGAGAACGCTTTCATTTGAATGAAATCCCGAGTATCTAGAGAG 850
      |||||||

```

297 pIleValGIuAsnArgrhEilEgIyMeIySserArgrGIyIeTyGIuT 314  
 851 CCCGAGAGGCGACCTCTTACCATGCTCATTTAACAATCGAGGCGCTTC 900  
 314 hPrIoLaGIyThrIleuTyrHisAlaHisLeuSprIleGIuAlaPhe 330  
 901 ACCATGAGCGGGAAGTGGCAAAATCAACAAAGCGCTGGCTTGAATT 950  
 331 ThMeTAsPArgrGIuValArgrLysIleLysGIuLysGIuLysPhe 347  
 951 TGGTGAGCTGTATACCGGTTTACGGCTTACGGCTGAGCTGGAATTGG 1000  
 347 eAlaGIuLeuValTyTrhGIyPheTrpHisSerProGIuLysGIuPheV 364  
 1001 TCCGCACTGATCGGCAAGTCCGAGGAGCGAGTGAAGGAAAGTGCAG 1050  
 364 aLArgrHisCysIleAlaLysSerGIuArgrValGIuGIyLysValGIu 380  
 1051 GTGTCCGCTCAAGGCGCAGGTATCATCTTCGGCGGAGTCCCACT 1100  
 381 ValSerValIleuLysGIuValGIuValIleuGIyArgrLysProLe 397  
 1101 GTCTCTACATAGAGAGGTGTGACATGAAGTGCAGGGTATATG 1150  
 397 uSerLeuTyTrhAsnGIuLysGIuLysValSerMeLAsnValGIuLysPhe 414  
 1151 AGCGCACTGATCGCAGCGGTTTCATCAATCAATTCCTGAGCTGAG 1200  
 414 LuProThrsPArgrGIuValrgrIleuValHisAlaHisLeuArgrLeuLys 430  
 1201 GAATATCATCTCTCCAGAGGAGTGCATCTGCCAA 1236  
 431 GIuTyTrhIsArgrLeuGIuInseryValIthrAlaLys 442

seq.name: /cgn2\_6/ptodata/2/pae/PCTUS.COMB.pep:PCT-US00-26524B-6180

seq-documentation block:  
 ; Sequence 6180, Application PC/TUS0026524B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birste et. al.  
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
 ; FILE REFERENCE: PA005PCT  
 ; CURRENT APPLICATION NUMBER: PCT/US00/26524B  
 ; CURRENT FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6180  
 ; LENGTH: 442  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (22)  
 ; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
 ; PCT-US00-26524B-6180

alignment-scores:  
 Quality: 2123.00 Length: 412  
 Ratio: 5.178 Gaps: 0  
 Percent Similarity: 99.515 Percent Identity: 99.272

alignment\_block:  
 US-09-775-693-1 x PCT-US00-26524B-6180 ..  
 Align seg 1/1 to: PCT-US00-26524B-6180 from: 1 to: 442

1 ATGTCCAGCAAGGCTCCGTGGTTCTGGCTACAGTGGCGGCTGGAGCAC 50  
 ||||||||||||||||||||||||||||||||||||||||||||||||

31 MeIserSerIySGIySerValIValLeuAlaTySerGIyGIyLeuAspTh 47  
 51 CTGTGATCCTCTGTGGGTGAAGGACAAAGCTATGACGTATGCCT 100  
 47 rSerCysIleLeuValItrPheLysGIuGIuGIuGIuTyTrhAspValIleAlaIat 64  
 101 ATCTGGCCCAACATTGGCCAGAGAGAACATTCTGAGAACACCCGAGAGAG 150  
 64 yLeuAlaHisnIleGIyGIuLysGIuAsPheGIuGIuAlaArgrLys 80  
 151 GCACGTGAACCTTGGGGCCCAAAAGGTGTTCATTGAGATGTCAAGCAGGA 200  
 81 AlaLeuLysLeuGIyAlaLysLysValPheIleGIuAsPArgrValSerArgr 97  
 201 GTTGTGGAGAGATTCACTGTGGCGCCCATCCAGTCCAGGCGCACTGTATG 250  
 97 upheValGIuGIuPheIleTrpProAlaIleGIuInserySerAlaLeuTyTr 114  
 251 AGGACCGCTACTCTGTGGGACCTCTTGGCAGGACCCCTGCATCGCCCGC 300  
 114 LuAsPArgrTyTrhLeuGIuGIuTrhSerLeuAlaArgrProCysIleAlaArgr 130  
 301 AAACAGTGAAGAAATCGCCAGCGGAGGGGCCAATGTGTGCCACGG 350  
 131 LysGIuValGIuIleAlaGIuArgrGIuGIuAlaLysTyTrhValSerHisGI 147  
 351 CGCCACAGCAAGAGGAGATCAGGTCCGGTTTGTAGCTCAGCTGTCT 400  
 147 yAlaTrhGIyLysGIuLysAsnSprGIuValArgrPheGIuLeuSerCysTyTrs 164  
 401 CACTGGCGCCCGCAGATAAGGTCAATTCCTCTCCCTGGAGAGTCAATTC 450  
 164 eTruAlaIleProGIuLysValIleAlaProItrPArgrMeTrProGIuPhe 180  
 451 TACACCGGTTCAAGGCGGCAATGACCTGATGAGTAGCCAAAGCAACA 500  
 181 TyTrhAsnArgrPheLysGIyArgrAsnSprLeuMeTrGIuTyTrhAlaLysGIuTh 197  
 501 CGGATATCCCATCCCGGTCACTCCCAAGAACCCGTGGAGCATGAGAGA 550  
 197 SGyIleProIleProValTrhProLysAsnProItrPserIleuAspIleu 214  
 551 ACCTCATGCACTACGTGAGAGGTGGAATCTTGAGAACCCCAACAAAC 600  
 214 snLeuMeLHisIleSerTyGIuAlaGIyIleLeuGIuAsnProLysAsn 230  
 601 CAAGCGCTCCAGGTCTCTACACGAAGACCCGAGACCCCAAGCCGCC 650  
 231 GIuAlaIleProProGIyLeuTyTrhLysTrhGIuAsnProAlaLysAlaPr 247  
 651 CAACACCCCTGACATCTCGAGATCGAGTTCAAAAAGGGTCCCTGTGA 700  
 247 oAsnTrhProAsnIleLeuGIuIleGIuPheLysLysGIyValProValL 264  
 701 AGGTGACCAAGCTCAAGATGGCACCAACCCAGACCTCTTGGAGCTTC 750  
 264 yValAlaTrhAsnValLysAsnGIyTrhTrhHisGIuTrhSerLeuGIuLeu 280  
 751 TTGATGTACCTGAGCAAGATCGCGGCAAGCATGGCTGGCGGTATTGA 800  
 281 PheMeTrTyLeuAsnGIuValAlaGIyLysHisGIyValGIyArgrLys 850  
 801 CATGTGAGAACCGCTTCATTGGAATGAAGTCCGAGGTATCTACAGAGA 850  
 297 pIleValGIuAsnArgrhEilEgIyMeIySserArgrGIyIeTyGIuT 314  
 851 CCCGAGAGGCGACCTCTTACCATGCTCATTTAACAATCGAGGCGCTTC 900  
 314 hPrIoLaGIyThrIleuTyrHisAlaHisLeuSprIleGIuAlaPhe 330  
 901 ACCATGAGCGGGAAGTGGCAAAATCAACAAAGCGCTGGCTTGAATT 950  
 331 ThMeTAsPArgrGIuValArgrLysIleLysGIuLysGIuLysPhe 347

```

951 TCCTGAGCTGTGTATACCGGTTTACGGCCTTACGCCCTGAGTGTGATTTG 1000
347 ealagluleuvaltyrthgltphetrphisserproglucylunphev 364
1001 TCCGACATGCATCCGCAAGTCCAGAGCAGAGGAGGAAAGGAAAGTGCAG 1050
364 alarghiscysillealalysseringluarvalgluglylvalgin 380
1051 GTTCCTGCTCTCAAGGCGCAGGTGTACATCTCCGCCGGAGTCCCACT 1100
381 valservalleuylsglylnvaltyrilleuglyargluserprole 397
1101 GTCTCTACATGAGAGAGCTGTGAGCATGAGAGTGCAGGAGTATATG 1150
397 uerleuylrasmelngluuvalsermetasnvalglnglylasprry 414
1151 AGCCACTGATGCCACCGGTTTCATCAATCATATTCCTCAGGCTGAG 1200
414 luprothraspalathrglypheelasnilleasnserleuArlguleys 430
1201 GAATATCATGCTCTCCAGAGCAAGTCACTGCCAAA 1236
431 Glutylhisarlgleninseryvalthralalys 442

seq_name: /cg02_6/plodata/2/paa/US099_COMB.pep:us-09-925-300-1594

seq_documentation_block:
/ Sequence 1594, Application US/09925300
/ GENERAL INFORMATION:
/ APPLICANT: Craig Rosen,
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA101
/ CURRENT APPLICATION NUMBER: US/09/925,300
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05988
/ PRIOR FILING DATE: 2000-03-08
/ NUMBER OF SEQ ID NOS: 1999-03-12
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1594
/ LENGTH: 442
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (22)
/ OTHER INFORMATION:
US-09-925-300-1594

```

alignment\_scores: xaa equals any of the naturally occurring L-amino acids

Quality: 2123.00 Length: 412  
 Ratio: 5.178 Gaps: 0  
 Percent Similarity: 99.515 Percent Identity: 99.272

alignment\_block:  
 US-09-775-693-1 x US-09-925-300-1594

Align seg 1/1 to: US-09-925-300-1594 from: 1 to: 442

```

1 ATGTCCAGCAAGGCTCGTGTGTGCTGCTACAGTGGCGGCTGAGCAC 50
31 MetSerSerlysglyservalvalleuvaltyrserglylyleuasp 47
51 CTGTGCATCTCTGCTGTGCTGAGCAAGCAAGCTATGAGTCATGCT 100
47 rsercysilleuvaltyrleuylsglylnvaltyrvalleuval 64
101 ATCTGCCACATTTGCCAGAGCAAGCACTTCAGAGAGCCAGGAAGAG 150

```

```

64 ylleuAlasnilleglylnlyslunaspheglunluAlarglyllys 80
151 GCACAGAGCTTTGGGGCCAAAAGCTGTTCATTTGAGAGATGTCAGCAGGA 200
81 AlaleuylsleuuglylvaltyrlyslvalpheelieglunaspvalserArg 97
201 GTTTGTGAGGAGCTTCATCTGGCCGGCCATCCAGTCCAGCGCAGCTATG 250
97 uphevalglunluuupheletrproAlalleglinserserAlaleuylrg 114
251 AGAGCCCTTACCTCTGGGACCTCTCTGCGCAAGCCCTGCATGCCCGC 300
114 lunspatglyrleuueuglythrserleuAlaApgrocyilleAlaArg 130
301 AAACAAGTGAAATCGCCCGAGGGAGGGGGCCAAATGATGTGCCACGG 350
131 LysgluvalgluileAlaAglargluilelylAlalystyrvalserHis 147
351 CGCCACAGCAAGAGGGAGCAAGATCAGTCCGGTTTGACCTGAGTCTACT 400
147 yAlathtglylsglyasnaspgluvalargpheeluileusercysTyr 164
401 CACTGGCCCCCAGATAAAGTCAATTCCTCTGAGAGATGCCGAATTC 450
164 erleuAlAproglinlelyvalleAlaAprotrpargmeproglu 180
451 TACAAACCGTTCAAGSGCCGCAATGACCTGATGAGTACCGCAAGCAACA 500
181 TyrasnArgpheylsglyarasnaspheumetgluylrAlalysglu 197
501 CGGATTCCTCCATCCGCTACATCCCAAGAACCCGTGAGCATGATGAGA 550
197 sglylleprolleprovalthrprolyasnprotrpsermetasp 214
551 ACCGATGACATCACTTCAAGAGCTGGAATCTGGGAACCCCAAGAAC 600
214 snleumethisleserlyrgluAlaglyleuuglunspProlyasn 230
601 CAAGCGCTCCAGTCTCTACACGAGAGCAAGCCAGCCAGCAAGCCCC 650
231 GluAlAprogluyleuyltyrthlysthglnaspProAlalysAl 247
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGGCTCCGTGGA 700
247 oasnrthproaspilleuugluileupheylslyslvalproval 264
701 AGGTACCAACGTCAAGATGACACCAACCCAGACCTCTTGAGAGTC 750
264 yslvalthrasnvallyslaspjlythrthslglnthserleuglu 280
751 TTCATGTACTCTGAAGAGTCCGGGCAAGCAATGCGTGGCGCTATTGA 800
281 PheuetrylleuasnluvalAlaAgllyslshgylvalglArglleas 297
801 CACGTGGAGAACCGCTTCATGCAATGAAGTCCCGAGGATGTACGAGA 850
297 pillevalgluasnargpheeliegllymetlysserarglylyleu 314
851 CCCACAGAGCACATCTTACCATGCTCATTTAGACATGAGGCTTC 900
314 hrproAlaAgllythrilleuutyrhisAlahislsuasplleglunAla 330
901 ACCATGAGACGGGAGTCCGCAAAATCAAAACAGGCTGCTGCTGAAAT 950
331 Thrmetaspargluvalarglysllelyslnglyleuglyleuyls 347
951 TCCTGAGCTGTGTATACCGGTTTACCGGCTTACAGTGGCGGCTGAGCAC 1000
347 ealagluleuvaltyrthgltphetrphisserproglucylunphev 364
1001 TCCGACATGCATCCGCAAGTCCAGAGCAGAGGAGGAAAGTGCAG 1050
364 alarghiscysillealalysseringluarvalgluglylvalgin 380

```

```

1051 GTGTCGCTCAAGAGCCAGGTGACATCCGCGCGAGTCCCACT 1100
      |||||||
391 ValSerValIleuLysGlyInAlaIleuLysGlyArgGluSerProIe 397
      |||||||
1101 GTCTCTACATGAGAGAGCTGTGAGCATGAACGTGCAGGGTGATATG 1150
      |||||||
1151 AGCCAACTGATGCCACCGGTTTCATCAACATCATTCCTCCGTGGTGAG 1200
      |||||||
414 LnpProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 430
      |||||||
1201 GAATATCATGCTCCAGAGCAAGGCACTGCCAAA 1236
      |||||||
431 GlutYrHisArgLeuGlnSerLysValThrAlaLys 442

seq_name: /cgn2_6/plodata/2/paa/us097_COMB.pep:us-09-760-475-2729

seq_documentation_block:
: Sequence 2729, Application US/09760475
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT49
: CURRENT APPLICATION NUMBER: US/09/760,475
: PRIORITY FILING DATE: 2001-01-16
: Prior application data removed - consult PAM or file wrapper
: NUMBER OF SEQ ID NOS: 4122
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2729
: LENGTH: 473
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (14)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (17)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-760-475-2729

alignment_scores:
      Quality: 2123.00      Length: 412
      Ratio: 5.178      Gaps: 0
Percent Similarity: 99.515      Percent Identity: 99.272

alignment_block:
US-09-775-693-1 x US-09-760-475-2729 ..

Align seg 1/1 to: US-09-760-475-2729 from: 1 to: 473

1 ATGTCCAGCAAGGCTCCGTGGTTCTTGCCCTACAGTGGCGGCTGGACAC 50
      |||||||
62 MetSerSerLysGlySerValIleuAlaIleuLysSerGlyLeuSerPth 78
      |||||||
51 CTCTGATCTCTGCTGGGTGAAGAAACAAGCTATGACGTACCTTGGCT 100
      |||||||
78 rSerCysIleLeuValIleuLysGlyGlnGlyTyrAspValIleLeuIat 95
      |||||||
101 ATCTGGCCACATTTGGCCAGAGAAAGACTTTCGAGAAAGCCAGAGAGAG 150
      |||||||
95 yIleuAlaAsnIleGlyGlnLysGlnAspPheGlnIleAlaArgLysLys 111
      |||||||
151 GCACCTGAAGCTTGGGGCCAAAAGTGTTCATGTGAGATGTGCACAGGGA 200
      |||||||
112 AlaLeuLysLeuGlyAlaLysLysValPheIleGlnLysPValSerArg 128
      |||||||
201 GTTGTGGAGAGATTTCATCTGCGCGCCATCCAGTCCAGCGCACTGATATG 250
      |||||||
128 uPheValIleuGlnIleuPheIleIleuProAlaIleGlnSerSerAlaLeuTyrG 145

```

```

251 AGAAGCCGCTACCTCTGGGCACTCTCTTGCACAGGCCCTGCATCCGCCG 300
      |||||||
145 LuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 161
      |||||||
301 AAACAAGTGGAAATCCGCCACGGGAGGGGAGGCAAGATATGTGCCACGG 350
      |||||||
162 LysGlnValGlnIleAlaGlnIleArgGlnGlyAlaLysTyrValSerHisG 178
      |||||||
351 CGCCACAGGAAGAGGGAGAGATCAGTCCGCTTTCAGCTCAGCTGCTACT 400
      |||||||
178 yAlaThrGlyLysGlyAsnAspGlnValArgPheGlnLeuSerCysTyrS 195
      |||||||
195 eIleuAlaProGlnIleLysValIleAlaProIleArgMetProGlnPhe 211
      |||||||
451 TACAACCGCGTTCAAGGGCGGCAATGACCTGTGTGAGATGACCAAGACA 500
      |||||||
212 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlnTyrAlaLysGlnH 228
      |||||||
501 CGGATTCCTCATCCCGGTCACTCCCAAGAACCCGTGAGCATGGATGAGA 550
      |||||||
228 SGlyIleProIleProValThrProLysAsnProIlePheSerMetAspGlu 245
      |||||||
551 ACCTCATGCACATACGCTACAGAGCTGGAAATCCTGGAGAACCCCAAGAAC 600
      |||||||
245 snLeuMetHisIleSerTyrGlnIleGlyIleLeuGlnAsnProLysAsn 261
      |||||||
601 CAAGCGCTCCAGGCTCTCTACAGAGAACCCAGAGCCAGCCAAAGCCCC 650
      |||||||
262 GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 278
      |||||||
651 CAACACCCCTGACATTTCTGAGATCGAGTCAAAAAGGGTCCCTGTGA 700
      |||||||
278 oAsnThrProAspIleLeuGlnIleGluPheLysGlyValIleProValL 295
      |||||||
701 AGGTACCAACAGCTCAAGATGGACACCCACAGACCTCCTGTGAGATC 750
      |||||||
295 yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGlnLeu 311
      |||||||
751 TTCATGTACCTGAACAGATCCGCGGCAAGCATGGCGTGGCCGTATTGA 800
      |||||||
312 PheMetTyrLeuAsnGlnValAlaIleGlyLysHisGlyValGlyArgIleAs 328
      |||||||
801 CATCGTGAAGAACCCGTTTCATTTGGAATGAATGCCGAGATATCTACAGA 850
      |||||||
328 pIleValGlnAsnArgPheIleGlyMetLysSerArgGlyIleTyrGlnT 345
      |||||||
851 CCCGACGAGCAGACATCCCTTACCATGCTCATTTAGACATGAGGCTTTC 900
      |||||||
345 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 361
      |||||||
901 ACCATGGACCGGGAAGTGGCAAAATCAACAAGAGCTTGGGCTTGAATT 950
      |||||||
362 ThrMetAspArgGlnValArgLysIleLysGlnGlyLeuLysPhe 378
      |||||||
951 TCGTGAAGCTGTGATACCGGTTTACGGCTTACGCTGAGTGTGATTTG 1000
      |||||||
378 eAlaGlnLeuValTyrThrGlyPheThrPheIleSerProGlnCysGlnPhe 395
      |||||||
1001 TCCGCACTGCATCGCCAAAGTCCAGAGAGAGAGTGAAGAGGAAGTGCAG 1050
      |||||||
1051 GTTCTCGTCCCTCAAGGGCCAGGTGTACATCTCGCGCGGAGATCCCACT 1100
      |||||||
412 ValSerValLeuLysGlyGlnValTyrIleLeuGlyArgGlnSerProIe 428
      |||||||
1101 GTCTCTCAATGAGAGAGCTGTGAGCATGAACGTGCAGGGGATTTATG 1150
      |||||||
428 uSerLeuTyrAsnGlnGlnLeuValSerMetAsnValGlnGlyAspTyrG 445

```

1151 AGCCACTGATGCGCCGGGTCATCAACATCAATTCCTCAGCTGAAG 1200  
 445 IuprotinaspalaThrGlyPheIleasnIleasnSerleuArgLeuLys 461  
 1201 GAATATCATGCTCCAGACAGCATGTCACGCCAAA 1236  
 462 GlutyrHisArgLeuGlnSerLysValThrAlaLys 473

seq\_name: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep:US-60-207-216-791

seq\_documentation\_block:

; Sequence 791, Application US/60207216  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Beasley, Ellen  
 ; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,  
 ; FILE REFERENCE: CLO00580  
 ; CURRENT APPLICATION NUMBER: US/60/207,216  
 ; CURRENT FILING DATE: 2000-05-26  
 ; NUMBER OF SEQ ID NOS: 898  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 791  
 ; LENGTH: 425  
 ; TYPE: PRT  
 ; ORGANISM: HUMAN  
 ; US-60-207-216-791

alignment\_scores:

Quality: 1631.00 Length: 412  
 Ratio: 4.660 Gaps: 2  
 Percent Similarity: 84.951 Percent Identity: 80.825

alignment\_block:

US-09-775-693-1 x US-60-207-216-791 ..

Align seg 1/1 to: US-60-207-216-791 from: 1 to: 425

1 ATGTCACGAAGAGCTCCGTTCTGCTACAGTGGCGCTGAGAC 50  
 50 MetSerSerLysGlySerValIleuAlaIleuArgLysGlyLeuAspTh 66  
 51 CTGCGTCACTCCCTGCTGCTGAGGAACAAGCTATAGCTCATTTGCTT 100  
 66 rSerCysIleuLeuTrpLeuLysGlnGlnGlyTyrAspValThrAla 83  
 101 ATCTGGCCACATTGGCCAGAGAGACTTCAGAGAGCCAGAGAGAG 150  
 83 yLLeuAlaAsnIleGlyIleLysGlnAspIheGlnAlaArgLysLys 99  
 151 GCACGTGAAGCTTGGGCCAAAAGGTGTCATTGAGGATGTCAAGAGGA 200  
 100 AlaProLysLeuGlyThrLysLysValPheThrGlnGlyValSerArgG 116  
 201 GTTGTGAGAGAGTTCATCTGGCGGCGCATCCAGTCCAGCGCATGTATG 250  
 116 upheValGlnIleuPheIleTrpProAlaIleGlnSerSerLeuTrg 133  
 251 AGAACCGCTACCTCTGGGCACTCTCTGCGAGAGCCCTGCATGCCGCG 300  
 133 IuaSpGlyTyrLeuLeuGlyThrSerHisThrArgProCysThrAlaArg 149  
 301 AAACAAGTGGAAATCGCCACGGGAGGGGCCCAAGTATGTGCCACGG 350  
 150 Lys..... 150  
 351 CGCCACAGAAAGGGAACGATCAGTCCGTTTTCAGCTCAGCTGACT 400  
 150 ..... 150  
 401 CACTGCGCCCCGAGATTAAGGTCAATTGCTCCCTGAGAGATGCTGAATTC 450  
 ..... 450

151 .....GlnIleuValIleAlaProTrpSerMetProGlnPhe 163  
 451 TACAAACGGTTCAAGGGCGCAATGACCTGATGAGTACGCAAGACAA 500  
 164 TyrAsnHisPheLysGlyCysSerAspLeuThrGlnGlyThrLysGlnI 180  
 501 GGGGATTCCTCCCGGCTCACTCCCAAGAACCCCTGGAGCATGGATGAGA 550  
 180 GcLylIleProIleProValThrProLysAsnProTrpAsnMetAspGln 197  
 551 ACCTCATGACATGACGTACGAGGCTGGAAATCCTGGAGAA...CCCAAGAA 599  
 197 snLeuMetHisIleSerGlyGlnAlaGlyIleLeuGlnAspProGlnGln 213  
 600 CCAAGGCGCTCCAGGTTCTTACAGGAAGCCAGAGCCCAAGCCAAAGCC 649  
 214 ProSerAlaSerArgSerLeuHisGlnAspProGly...LeuAlaLysAla 230  
 650 CCAACACCCCTGACATCTCGAGATCGAGTTCAAAAAGGGGTCCGTG 699  
 230 toAsnThrProAspIleLeuGlnIleGlnIleuLysLysGlyValIleLeu 246  
 700 AAGGTGACCAAGTCAAGATGGCACCCACCCAGACCTCTTGAGAGT 749  
 247 LysValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuAla 263  
 750 CTTCATGACCTGACAGAGTCCCGGGCAAGCATGGCGTGGGCGGATTC 799  
 263 upHeMetGlyLeuAsnGlnValThrGlyLysHisCysValGlyArgIle 280  
 800 ACATGCTGGAGAACCGCTTCATGGAATGAAGTCCCGAGTATCTACAG 849  
 280 spIleValGlnAsnCysPheIleGlyMetLysSerArgGlyIleGlyGln 296  
 850 ACCCCAGCAGCAACATCTTACATGCTCATGTTGACATCGAGGCTT 899  
 297 ThrProAlaGlyThrIleLeuGlyThrHisAlaHisLeuAspIleGlnThr 313  
 900 CACCATGGACGGGGAAGTGGGCAAAATCAAAACAAGGCTGGGCTTAAT 949  
 313 ethrMetAspTrpGlnValArgLysIleLysGlnGlyLeuLysC 330  
 950 TTGCTAGCTGCTGTATACCGGCTTACCGGCTTACCGGCTTACCGGCT 999  
 330 ysaIaGlnLeuValGlyThrGlnPheTrpHisSerProLysGlnPhe 346  
 1000 GTCCGCGCATGTCATGCCCAAGTCCCAAGAGCAGTGGAGGAAGTGA 1049  
 347 ValArgHisSerIleAlaLysSerGlnGlnArgValGlnArgLysVal 363  
 1050 GGTGTCGCTCTCAAGGGCCAGGTGACATCTCCGGCCGGGAGTCCCAAC 1099  
 363 nValSerLeuPheLysAspGlnValIleGlyLeuGlyTrpGlnSerPro 380  
 1100 TGTCTCTTACAATGAGAGCTGTGAGCATGACATGACGAGGTGATAT 1149  
 380 rSerLeuTrpLysGlnGlnIleuValSerMetAsnValHisGlyLysPyr 396  
 1150 GAGCCAGCTAGTCCACCGGTTTCATCAACATCAATTCCTCAGCTGAA 1199  
 397 GluProIleAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeu 413  
 1200 GGAATATCATGCTTCCAGAGCAAGTCACTGCC 1233  
 413 GcLutyrHisHisLeuGlnSerLysValThrAla 424

seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep:PCT-US01-08631-34811

seq\_documentation\_block:

; Sequence 34811, Application PC/TUS0108631  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

```

FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT-US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 34811
LENGTH: 433
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (112)..(157)
OTHER INFORMATION: Argininosuccinate synthase protein domain identified by
OTHER INFORMATION: EMATRIX, accession number BL00564C, p-value=1.000e-40, raw score
OTHER INFORMATION: 25.47
NAME/KEY: DOMAIN
LOCATION: (8)..(315)
OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
OTHER INFORMATION: name Argininosucc_synth, E-value=5.5e-151, Pfam score of 481.4
NAME/KEY: misc.feature
LOCATION: (1)..(433)
OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-34811

```

## alignment\_scores:

```

Quality: 1603.50      Length: 428
Ratio: 4.310          Gaps: 15
Percent Similarity: 86.916      Percent Identity: 80.140

```

## alignment\_block:

```
us-09-775-693-1 x PCT-US01-08631-34811 ..
```

```
Align seg 1/1 to: PCT-US01-08631-34811 from: 1 to: 433
```

```

1 ATGTCCAGCAAAAGGCTCCGTGCTCTAGCTACAGTGGCGGCTGGACAC 50
1 MetSerSerIysGlySerValValLeuAlaIyrSerGlyGlyLeuAspTh 17
51 CTGTGCATCTCTGTGTGCTGAAGGAACAAGCTAGCTATGCTT 100
17 rSerCysIleLeuValIrrPheLysGlyGlyIrrPheValIrrPheVal 34
101 ATCTGCCCAACTTGGCCAGGAAGAACTTCAGAGAGCCAGGAAGAG 150
34 yLeuAlaAsnIleGlyIrrPheLysGlyGlyIrrPheValIrrPheVal 50
151 GCACGTGAAGCTTGGCGGCAAAAGGCTTCATGAGATGTCCAGAGGA 200
51 AlaLeuIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 67
201 GTTGTGAGAGGATTCATCTGCGCCAGCTCAGTCCAGCCACTGTATG 250
67 uPheValIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 84
251 AGGACCGCTACTCTCTGGGCACTCTCTTGGCAGGCGCTGCATGCCGC 300
84 LuAspArgIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 100
301 AAACAGGTGAATGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 350
101 LysGlyValIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 117
351 CGCCACAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 400
117 yAlaThrGlyLysGlyAsnAspGlyValIrrPheLysGlyGlyIrrPhe 134
401 CACTGGCCCCCAGATAAAGTCAATGCTCCCTGAGAGATGCCGAATTC 450

```

```

134 eLeuAlaProGlnIleLysValIrrPheValIrrPheValIrrPheVal 150
451 TACAAAGGCTTCAAGGCGCAATGACCTGATGAGTACGCAAGGAGAGA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGlyTyrAlaLysGln 167
501 CGGATTTCCATCCCGGTCTACTCCCAAGAACCCGTGAGATGATGAGA 550
167 sGlyIrrPheProIrrPheValIrrPheValIrrPheValIrrPheVal 184
551 ACCGATGCAATCAGCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
184 sLeuMetHisIleSerTyrGlyAlaIrrPheValIrrPheValIrrPheVal 200
601 CAAGCGCTCCAGGCTCTACAGGAAGGAGGAGGAGGAGGAGGAGGAGGAG 650
201 GlnAlaIrrPheProGlyLeuIrrPheValIrrPheValIrrPheVal 217
651 CAACACCCCTGACATTCGAGATCGAGTTCGAGGAGGAGGAGGAGGAGGAG 700
217 oAsnThrProAspIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 234
701 AG...GTGACCAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 747
234 LuGlyGlyProIrrPheLysAspGlyIrrPheValIrrPheValIrrPheVal 250
748 CTCTCATGTACTGAGGAAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794
251 LeuPheMetIrrPheLysGlyGlyIrrPheValIrrPheValIrrPheVal 267
795 TATTGACATGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 841
267 rLeuThrSerTrpIrrPheLysAsnArgPheHisIrrPheValIrrPheVal 284
842 TCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 888
284 LeuAlaIrrPheProGlnIrrPheValIrrPheValIrrPheValIrrPheVal 300
889 ATCGAGGCTTCAACATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
301 IleGlnAlaIrrPheMetIrrPheValIrrPheValIrrPheValIrrPheVal 317
936 CTGTGGCTTGAATTTGCTGAGCTG...GTGTATACCGGTTTACGGCTTA 962
317 yLeuGlyPheGlyIrrPheValIrrPheValIrrPheValIrrPheVal 334
963 GCCGTGAGTGAATTTGCTGAGCTG...GTGTATACCGGTTTACGGCTTA 1029
334 IrrPheLysGlyGlyIrrPheValIrrPheValIrrPheValIrrPheVal 350
1030 CGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1073
351 ArgValIrrPheLysSerAlaGlyValIrrPheValIrrPheValIrrPheVal 367
1074 GTACATCTCTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1120
367 sThrSerLeuGlyIrrPheValIrrPheValIrrPheValIrrPheVal 384
1121 TGTGAGCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1164
384 euValIrrPheLysGlyIrrPheValIrrPheValIrrPheValIrrPheVal 400
1165 ACCGAGTTCATCAACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1205
401 ProGlyIrrPheLysGlyIrrPheValIrrPheValIrrPheValIrrPheVal 416
1206 TCATGCTCTC...CAGAGCAAGGCTACTGCCA 1236
416 rHisArgLeuProArgAlaLysValIrrPheValIrrPheValIrrPheVal 427

```

```
seq_name: /cgn2_6/ptdata/2/paa/PCTUS_COMB.pep:PCT-US01-08631-49199
```

```
seq_documentation_block:
```



```

; Sequence 49199, Application PC/TU050108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49199
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (7)..(44)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by
; OTHER INFORMATION: eMARTIX, accession number BL00564A, p-value=1.231e-27, raw score
; OTHER INFORMATION: 19.93
; NAME/KEY: DOMAIN
; LOCATION: (8)..(333)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
; OTHER INFORMATION: name Arginosuc_synth, E-value=7.7e-145, Pfam score of 461.9
PCT-US01-08631-49199

```

```

alignment_scores:
  Quality: 1431.50      Length: 407
  Ratio: 4.618          Gaps: 2
  Percent Similarity: 76.167  Percent Identity: 70.762

```

```

alignment_block:
US-09-775-693-1 x PCT-US01-08631-49199 ..

```

```

Align seg 1/1 to: PCT-US01-08631-49199 from: 1 to: 385

```

```

1 ATGTCCAGCAAGAGCTCCGTGCTTCTGCGCTACAGTGGCGGCTGACAC 50
|||||
1 MetSerSerLysGlySerMetValLeuAlaTyrSerGlyLeuAlaSnTh 17
51 CTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100
|||||
17 rSerGylleLeuValThrLeuLysGluInGlyTyrAspValThrAlat 34
101 ATCTGGCCACATTTGGCCAGAGAGACTTGCAGAGACCAGAGAGAGAG 150
|||||
34 yLeuAlaSnThrGlyInGlyGluAspPheGluInGluAlaArgLysLys 50
151 GCACCTAAGCTTGGGGCAAAAAGGTTCATTGAGAGATGTCACAGAGGA 200
|||||
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerArgL 67
201 GTTTGGAGAGAGATTTCATCTGGCGGCGCATTCAGTCCAGCGCATGTAT 250
|||||
67 urPheValIleGluInPheIleThrPalaIleIleGlnSerSerAlaLeuTyrG 84
251 AGGACGCTACCTCTCTGGGACCTCTCTTGCAGGCGCTGCATCGCCGCG 300
|||||
84 LuAspHisTyrLeuLeuGlyThrSerLeuThrArgProCysIleAlaHis 100
301 AAACAGTGAATCGCCAGCGGGAGGGGCGCAAGTATGTCTCCACGCG 350
|||||
101 LysLysValGluThrValGlnArgValGlyAlaLysTyrGluGlyProTr 117
351 CGGCACAGAGAAAGGAGACGATCAGTCCGTTTGAGTGCACCTGCTACT 400
|||||
117 pArg..... 118
401 CACTGGCCCCCAGATAAAGTATGTCCTCTGAGAGATGCTGAATTC 450

```

```

|||||
119 .....AlaProThrArgMetProGluPhe 126
451 TACAACCGGTTCAGAGCGCCGATGACCTGATGAGATGACGAAGACACA 500
|||||
127 TyrAsnThrPheLysGlyArgSerAspLeuMetGluTyrAlaLysGlnH 143
501 CGGATTCCTCCATCCCGGCTACCTCCCAAGAACCCCTGAGCATGATAGA 550
|||||
143 SGIYleProIleLeuValThrProLysAsnLeuTrpSerIleAspLysA 160
551 ACCTCATGACATCAGTACGAGGCTGGAATCTGTGGAACCCCAAGAAC 600
|||||
160 snLeuMetHisIleSerTyrLysAlaGlyIleLeuGluAsnProLysAsn 176
601 CAAGCGCTCCAGGTCTCTACAGCAAGACCCAGACCCAGCAAGCCGCC 650
|||||
177 GluValProProGlyLeuTyrThrLysIleGlnAspProAlaLysAlaPr 193
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAAGGGGCTCCGTGTA 700
|||||
193 AsnThrProAspIleLeuLysIleGlnPheLysLysGlyValProValA 210
701 AGGTGACCAACGTCAGAGATGGACCAACCCAGACCTCTTGAGCTC 750
|||||
210 rGValThrSerValLysAspGlyThrThrHisGlnThrSerLeuGluLeu 226
751 TTCATGTACTGCAACGAAGTCCGGGCAAGATGCGTGGCGCTATTGA 800
|||||
227 PheMetTyrLeuAsnGluValAlaGlyLysHisGlyValGlyTyrIleAl 243
801 CATCGTGAGAACCGCTTCATTGGAATGAAGTCCCGAGTATCTACGAGA 850
|||||
243 AlIleValGluAsn..... 247
851 CCCAGACGAGCACATCCCTTACCATGCTCATTTAGACATCGAGGCTTTC 900
247 ..... 247
901 ACCATGACCGCGGAAGTCCGCAAAATCAACAACAGCGCTTGGAATT 950
247 ..... 247
951 TGCTGAGCTGCTATACCGGTTACGGCTAGCCCTGATGTAATTG 1000
|||||
248 .....IlePheThrArgGlnArgProSerProGluCysLuePheV 261
1001 TCCGCCACATGCTGCAAGTCCAGAGAGCGAGTGAAGGAAAGTGCAG 1050
|||||
261 AlArgProCysIleAlaLysSerGlnGluProAlaGluGlyLysValGln 277
1051 GTGTCCGCTCCAGAGCGGCTGATCATCTCGCGCGGAGTCCCATCT 1100
|||||
278 ValProValIleuLysGlyGlnValTyrIleLeuGlyTyrPgluSerProLe 294
1101 GTCTCTACATAGAGAGCTGTGAGCATGAACGTGACAGGCTGATATTG 1150
|||||
294 uSerLeuTyrAsnGluGluLeuMetSerValAsnValGlnGlyAspTyrG 311
1151 AGCCAACTGATGACACCGGCTTCATCAACATCAATTCCCTAGGCTGAG 1200
|||||
311 LuProIleAspAspThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 327
1201 GAATTCATCTCTCCACAGC 1221
328 GluTyrHisCysLeuGlnSer 334
seq_name: /cgn2_6/plodata/2/paa/US60_COMB.pep:US-60-243-468-1235
seq_documentation_block:
; Sequence 1235, Application US/60243468
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen

```



```

: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
: FILE REFERENCE: AND USES THEREOF
: FILE REFERENCE: CL000929
: CURRENT APPLICATION NUMBER: US/60/243,468
: CURRENT FILING DATE: 2000-10-27
: NUMBER OF SEQ ID NOS: 2121
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1235
: LENGTH: 317
: TYPE: PRT
: ORGANISM: HUMAN
: US-60-243-468-1235

```

```

alignment_scores:
  Quality: 1390.00      Length: 342
  Ratio: 4.680          Gaps: 4
  Percent Similarity: 86.842  Percent Identity: 81.287

```

## alignment\_block:

US-09-775-693-1 x US-60-243-468-1235 ..

Align seg 1/1 to: US-60-243-468-1235 from: 1 to: 317

```

1 ATGTCAGCAAGGCTCCG...GTTCGGCTACAGTGGCGCTGGA 47
1 MetSerSerLysGlySerMetArgValLeuAlaTyrGlyGlyLeuAs 17
48 CACCTCGTCATCTCTGTCGTGCTGAAGACAGCAGTACGCTATTG 97
17 pThrSerCysIleLeuValIleValIleValIleValIleValIle 34
98 CCTATGCGCAACATGTCGCGCAGAGAGAGACTTGAGAACCCAGAG 147
34 L TyrLeuAlaAsnThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 50
148 AAGGCCTAGAGCTTGGCGCAAAAGGTTCATTGAGATGTCAGAG 197
51 LysAlaLeuLysLeuGlyAlaLysLysValPheIleGlyAspValIle 67
198 GGAATTGTGAGAGAGTTCATCTGCGCGGCATCCAGTCCAGCAGCT 247
67 gLupPheValIleGluIupPheIleThrProAlaIleGlnSerThrLeu 84
248 ATGAGAGCCGCTACTCTCTGCGGACCTCTTGGCCAGGCCCTGCA 297
84 TyrLupAspArgTyrLeuLysGlyThrPheLeuAlaArgProCysIle 100
298 CGCAACAGTGAATCGCCCGCGGAGGAGGCGCAAGTATGTCGCA 347
101 SerLysGlnValIleLeuAlaGlnGlnGlnGlnGlnGlnGlnGln 114
348 CGCGCGCACAGAAAGGAGCAATCAGTCCGTTGAGCTCAGCTGCT 397
114 gGlyAlaThrGlyGlu.....TyrGlnValArgPheGlnLeuThrP 129
398 ACTCAGTGGCGCCCGCAGATTAAGTATGCTCTCTGCGGAGATGCT 447
129 TyrSerLeuValProGlnIleLysGlyThrAlaProThrPargMet 145
448 TTTCACACAGCGTTCAAGGCGCCAGTGAAGTATGAGTACCAAGA 497
146 PheTyrAsnArgPheLysGlyArgAsnAspLeuMetGlyTyrAla 162
498 ACACGGGATTCCATCCGGTCACTCCCAAGAACCCGTGAGCATGAG 547
162 nHisGlyIleProIleProValThrProLysAsnProThrPargMet 179
548 AGAACCTCAGCAGATCAGTACGAGGCTGAGATCTGAGGAGAACCC 597
179 LupsnLeuMetHisIleSerTyrGluAlaGlyIleLeuGlnAsnPro 195

```

```

598 AACCAAGCGCTCCAGGCTCTCTACAGCAAGACCCAGACCCCAAG 647
196 AsnGlnValProProAsp..... 201
648 CCCCAACACCCCTGACATTCCTCGAGATCGAGTTCAAAAAGGG 697
202 .....ArgValLysGlyValPro 209
698 TGAAGTGACCAAGCTCAGATGACACACACACAGACCTCTTGAG 747
209 allLysValThrAsnValAlaAsnAspLysSerThrHisProThr 225
748 CTCTTCATGTAACCTGAAGCAAGTCCGCGGCAAGCATGCGCGCT 797
226 PhePheMetLysLeuAsnGlnValAlaAspLysHisGlyValIle 242
798 TGACATGCTGAGAACCGCTTCATTTGATGAATGAGTCCGAGTAT 847
242 rAspIleValGlnAsnArgPheThrGlyMetLysSerArgGlyIle 259
848 AGACCCGACGACGACCATCTTACCATGCTCATTTAGACATGAGCC 897
259 IupThrProAlaGlyThrIleLeuTyrHisThrHisLeuAspIle 275
898 TTCACCATGACCGGAGAGTCCGCAAAATCAACACAGCTGGCTT 947
276 PheThrMetAspTrpGluValCysLysIleLysGlnIleLeuGly 292
948 ATTGTCGAGCTGCTGATACCGGTTACGCGCTAGCGCTGAGTGA 997
292 sPheAlaIleLeuValTyrThrAlaPheThrPHisSerProGlu 309
998 TTGTCCGCACTGCATCGCCCAAGTCC 1023
309 heValArgHisCysIleAlaLysSer 317

```

seq\_name: /cgn2\_6/ptodata/2/paa/US60\_COMB.dep:US-60-212-664-486

## seq\_documentation\_block:

```

: Sequence 486, Application US/60212664
: GENERAL INFORMATION:
: APPLICANT: Ladunga, Steve
: APPLICANT: Spier, Gene
: APPLICANT: Greenberg, Simon
: APPLICANT: Rabkin, Steven
: APPLICANT: Wang, Yu
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
: FILE REFERENCE: CL000687
: CURRENT APPLICATION NUMBER: US/60/212,664
: CURRENT FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 636
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 486
: LENGTH: 347
: TYPE: PRT
: ORGANISM: HUMAN
: US-60-212-664-486

```

```

alignment_scores:
  Quality: 1350.50      Length: 347
  Ratio: 4.502          Gaps: 3
  Percent Similarity: 86.455  Percent Identity: 78.386

```

## alignment\_block:

US-09-775-693-1 x US-60-212-664-486 ..

Align seg 1/1 to: US-60-212-664-486 from: 1 to: 347

```

1 ATGTCAGCAAGGCTCCG...GTTCGGCTACAGTGGCGCTGAGAC 50
|||||

```

```

1 MetSerSerLysGlySerIleValIleAlaCysSerGlySerLeuAspThr 17
51 CTGTGATCCTGCTGTGGCTGAGAACAGAGCTATGAGCTATTGCTT 100
17 rPheCysIleLeuValIleThrLeuLysGluGlnGlyTyrAspValIleAla 34
101 ATCTGGCAACATTTGGCCAGAGAAAGACTTCGAGAGAGAGAGAG 150
34 yIleuAlaSerIleGlyGlnGlyAspPheGlnIleAlaArgLysLys 50
151 GCACCTGAGCTTGGGCGCAAAAGGTTCATTGAGATGTCAGCAGGA 200
51 AlaLeuLysProGlyAlaLysLysMetLeuIleGluAspValAsnArg 67
201 GTTTGTGAGAGAGTTCATCTGGCGGCATCCAGTCCAGCAGCTATG 250
67 upheValIleGluIlePheIleThrProAlaMetGlnSerThrLeuTyrG 84
251 AGGACCGCTACTCTCTGGGACCTCTCTTCCAGGCTTCATGCGCCG 300
84 IuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleThrHis 100
301 AAACAGTGAATGCGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
101 LysGlnValIleValAlaGlnIleArgGluGlyAlaLysTyrValSerHisG 117
351 CGCCACAGGAAGAGGAGACGATCAGCGCTTGTGAGTCACCTGCTACT 400
117 yThrThrGlyLysGlyAsnAspGln..... 125
401 CACTGGCCCCCAGATAAAGTTCATTGCTCCTCGAGAGATGCTGAATTC 450
125 ..... 125
451 TACAACCGGTTCAAGGCGCCGCAATGACCTGATGAGTACGCAAGACA 500
126 .....GlyHisSerAspLeuThrGluTyrAlaLysGlnArg 137
501 CGGAGATTCCTCCATCCGGTCTACTCCCAAGAACCTGTGAGCATGATGAGA 550
137 gelyIleProThrProValThrProLysAsnProTrpSerMetAspGlu 154
551 ACCGTATGACATCAGTCAGAGGCTGAGATCTTGAGAACCCCAAGAAC 600
154 sIleuMetHisIleSerCysGlnAlaGlyIleLeuGlnAsnProGluAsn 170
601 CAAGGCGCTCCAGTCTCTACAGCAAGACCCAGACCCAGCAAAAGCCC 650
171 GlnAlaProProAlaLeuTyrThrLysThrGlnAspProAlaLysThrPr 187
651 CAACACCCCTGACATTCGAGATCGAGTCAAAAAGGGGTCCCTGTGA 700
187 GAsnThrProAspIleLeuGlnIleGluPheLysLysValIleProVal 204
701 AGGTACCAAGCTCAAGATGGACACCCACCCAGCTCCTTGAGAGTC 750
204 yValIleThrAsnValLysAspSerAlaThrHisGlnThrSerLeuGluLeu 220
751 TTCATGATCTGAAACGAGTGGCGGCAAGCATGGCGTGGCGCTGATGGA 800
221 PheMetTyrLeuAsnGluValAlaGlyLysHisSerValIleAsnIleAl 237
801 CATCTGGAGAAACGCTTCATTGGAAGTCCCGAGATTCACAGAGA 850
237 aIleValIleGluAsnArgPheIleGlyThrLysSerArgLysIleTyrArg 254
851 CCCCAGCAGGACCATCTTACATGCTGATTAGACATGAGAGCCTTC 900
254 IeProAlaGlyThrIleLeuCysTyrAlaHisLysAspIleGluAlaPhe 270
901 ACCATGACCGGGAAGTCCCAAAATCAACAGAGCTGGGCTTGAATTT 950
271 ThrMetTyrArgGluValAlaHisLysIleLysGlnLysArgGlyLeuLysPh 287

```

```

951 TCTGAGCTGTGTATACGGGTTTAC.....GGCCAGCCCTGAGT 991
287 eAlaGluLeuValIleThrThrGly..PheThrLeuSerGlyThrAlaLeuSer 303
992 GTGATTTGTCCCGACCTGCTATCCGCAAGTCCCAAGAGC 1030
304 ValAsnLeuSerAlaThrAlaSerProSerProArgSer 316

```

seq\_name: /cgn2\_6/prodata/2/paa/US60\_COMB pep:US-60-216-770-97

seq\_documentation\_block:

```

; Sequence 97 Application US/60216770
; GENERAL INFORMATION:
; APPLICANT: Ladunga, Steven
; APPLICANT: Spier, Eugene
; APPLICANT: Greenberg, Simon
; APPLICANT: Rabkin, Steven
; APPLICANT: Wang, Yu
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: C1000727
; CURRENT APPLICATION NUMBER: US/60/216,770
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 97
; TYPE: PRT
; ORGANISM: HUMAN
US-60-216-770-97

```

alignment\_scores:

Quality	Ratio	Length	Gaps
1350.50	4.502	347	3
Percent Similarity: 86.455		Percent Identity: 78.386	

alignment\_block:

US-09-775-693-1 x US-60-216-770-97 ..

Align seg 1/1 to: US-60-216-770-97 from: 1 to: 347

```

1 ATGTCCAGCAAGGCTCCGTGGTTTGGCTTACAGTGGCGCTTGACAC 50
17 rPheCysIleLeuValIleThrLeuLysGluGlnGlyTyrAspValIleAla 34
51 CTGTGATCCTGCTGTGGCTGAGAACAGAGCTATGAGCTATTGCTT 100
17 rPheCysIleLeuValIleThrLeuLysGluGlnGlyTyrAspValIleAla 34
101 ATCTGGCAACATTTGGCCAGAGAAAGACTTCGAGAGAGAGAGAG 150
34 yIleuAlaSerIleGlyGlnGlyAspPheGlnIleAlaArgLysLys 50
151 GCACCTGAGCTTGGGCGCAAAAGGTTCATTGAGATGTCAGCAGGA 200
51 AlaLeuLysProGlyAlaLysLysMetLeuIleGluAspValAsnArg 67
201 GTTTGTGAGAGAGTTCATCTGGCGGCATCCAGTCCAGCAGCTATG 250
67 upheValIleGluIlePheIleThrProAlaMetGlnSerThrLeuTyrG 84
251 AGGACCGCTACTCTCTGGGACCTCTCTTCCAGGCTTCATGCGCCG 300
84 IuAspArgTyrLeuLeuGlyThrSerLeuAlaArgProCysIleThrHis 100
301 AAACAGTGAATGCGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
101 LysGlnValIleValAlaGlnIleArgGluGlyAlaLysTyrValSerHisG 117

```

```

351 CGCCACAGAAAGGGAACGATCAGTCGGGTTGAGCTAGCTGCTACT 400
    |||:|||||
117 yThrThrclYlYsglyAsnAspGln..... 125
401 CAGTGGCCCCCAGATTAAGTTCATGTGCTCCCTGGAGGATGCTGATTC 450
125 ..... 125
451 TCACACCGGTTCAAGGCCGCAATGACCTGATGAGTACGCAAGCAACA 500
    |||:|||||
126 ..... GlyHisSerAspLeuThrGlnAlaYlYsglnAla 137
501 CGGAGTTCCTCCATCCGGTCACTCCACAGACCCGCGAGCATGATGAGA 550
    |||:|||||
137 gGlyLeuProThrProValThrProLYsAsnProTTrpSerMetAspGluA 154
551 ACCTCATGACATCAGCTACGAGGCTGAGTCTGAGAACCCCAAGAAC 600
    |||:|||||
154 snLeuMetHisIleSerCysGluAlaGlyIleLeuGluAsnProGluAsn 170
601 CAAGCCCTCCAGGCTCTACACGAAAGCCAGACCCAGCCAAAGCCCC 650
    |||:|||||
171 GlnAlaProProAlaLeuThrThrIlystrGlnAspProAlaIlystrPr 187
651 CAACACCCCTGACATTCCTGAGATGAGTTCAAAAAGGCGTCCCTGGA 700
    |||:|||||
187 oAsnThrProAspIleLeuGlnIleGlnPheLYsGlyValProValL 204
701 AGGTGACCAACGTCAGAGTGGACACCCACACGACTCCTTGGAGCTC 750
    |||:|||||
204 yValThrAsnValLYsAspSerAlaThrHisGlnThrSerLeuGlnLeu 220
751 TTCATGACCTGACAGAGTGGCGGCAAGCATGGCGTGGCGGTATTA 800
    |||:|||||
221 PheMetIlyLeuAsnGlnValAlaGlyLYsHisSerValGlyAsnIleAl 237
801 CATGTGGAGAACCGCTTCATTTGGAATGAGTCCGAGGATCTACAGAA 850
    |||:|||||
237 aIleValGluAsnArPheIleGlyThrLYsSerArgGlyIleIlyArgA 254
851 CCCGAGGAGCACCATTTCATTCATGCTATTAAGATCGAGGCGCTTC 900
    |||:|||||
254 laproAlaGlyThrIleLeuGlySTyAlaHisIleAspIleGluAlaPhe 270
901 ACCATGACCGCGGAGTGGCAAAATCAACAGGCGCTGGCTTGAAT 950
    |||:|||||
271 ThrMetIlyArgGluValHisLYsIleLYsGlnCysArgGlyLeuLYsPh 287
951 TGCTGAGTGGGTATACCGGTTTAC.....GGCCTAGCCCTGAGT 991
    |||:|||||
287 eAlaGluLeuValIlyThrngLY.PheTrpLeuSerGlyThrAlaLeuSer 303
992 GTGAATTTGTCGGCCTGATCGCCAGATGCCAGAGAC 1030
    |||:|||||
304 ValAsnLeuSerAlaThrAlaSerProSerProArgSer 316
seq_name: /cgn2_6/ptodata/2/paa/PCTus_COMB.pep:PCT-US01-08631-49202
seq_documentation_block:
; Sequence 49202, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49202

```

```

; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (23)..(60)
; OTHER INFORMATION: Argininosuccinate synthase protein domain identified by
; OTHER INFORMATION: EMATRIX, accession number BL00564A, p-value=2.800e-37, raw sco
; OTHER INFORMATION: 19.93
; NAME/KEY: DOMAIN
; LOCATION: (24)..(272)
; OTHER INFORMATION: Argininosuccinate synthase domain identified by Pfam, accession
; OTHER INFORMATION: name Argininosuc-synth, E-value=6.4e-68, Pfam score of 217.6
; NAME/KEY: misc_feature
; LOCATION: (1)..(487)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-49202

```

```

alignment_scores:
    Quality: 1314.00      Length: 445
    Ratio: 3.670          Gaps: 25
    Percent Similarity: 80.449      Percent Identity: 69.438

```

alignment\_block:

US-09-775-693-1 x PCT-US01-08631-49202 ..

Align seg 1/1 to: PCT-US01-08631-49202 from: 1 to: 487

```

1 ATGTCCAGCAAAAGGTCCTCGGTCTGCGCTACAGTGGCGGCTGAGAC 50
    |||:|||||
17 LeuSerIlysglySerValValIleAlaIlystrGlyLeuAspPh 33
51 CTCGTGATCCTCGTGTGCTGAGAGACAGGCTATGACGTCATTCGCT 100
    |||:|||||
33 rSerCylIleLeuValTrpLeuLYsGlnGlnIlyTrpAspValIleAla 50
101 ATGTGGCCAACTTGGCGCAGAGAGACTTCGAGAGACCGAGGAAG 150
    |||:|||||
50 yIleuAlaAsnIleGlyGlnLYsGlnAspPheGlnGluAlaArgLYs 66
151 GCACGTGAAGCTTGGGGCCAAAGAGTGTTCATTGAGATCTCAGCAGGA 200
    |||:|||||
67 AlaLeuLYsLeuGlyAlaIlyLYsValPheIleGluAspValSerArg 83
201 GTTGTGAGAGATTTCATCTGCGCGGCATTCGATCCAGTCCAGGCA...CTG 247
    |||:|||||
83 uPheValGlnGluPheIleTrpProAlaIleGlnValGlnAlaHisLeu 100
248 ATGAGAGACCGGTACCTCCTGGGC...ACCTCTCTTCCAGGCGCTGCATC 294
    |||:|||||
100 yGluAspArgGlyTrpLeuProGlyHisLeuSerLeuProGlyProCysIle 116
295 GCCCGCAAAACAAGTG...GAATGCGCCGAGGGAGGGGCGCAAGTAT.. 339
    |||:|||||
117 AlaArgLYsGlnValGlyAsnSerProSerIlyArgIlyAlaIlySTyVa 133
340 .GTGTCCACAGGCGGCACAGAAAGGGAAGCATGAGTCCGCTTAGC 388
    |||:|||||
133 IValProThrGlyAlaThrngLYsGlyAsnGluValArgPheGln 150
389 TCAGTGTACTACTGTGCGCCCGCCAGATTAAGATCATTCCTCGTGAAG 438
    |||:|||||
150 euserCylSerIleuAlaProGlnIleLYsValIleAlaProTrpArg 166
439 ATGCCGTGATTCACACCGGTTCAAGGCGCAGTATGATGAGTA 488
    |||:|||||
167 MetProGluPheTrpAsnArgPheLYsGlyArgAsnAspLeuMetGluTY 183
489 CGCAAGCAACAC...GGGATTCACATCCCGGTGACTCCCAAGAACCGGT 535
    |||:|||||
183 rGlyLYsAlaThrLeuGlyPheProIleProValThrProLYsAsnProT 200

```

```

536 GAGACATGATGAGAACCTCATGACATGACATGAGAGCTGGAATCCG 585
200 rpsrMetAspGluAsnLeuMetHisIleSerTyrGlnAlaIleu 216
586 GAGACGCGGAGAACGACGCGCTCCAGCTCTCTACAGACGACCCAGA 635
217 GlnAspTyrAsnGlnAlaProProGlyLeuTyrThrIleGlnI 233
636 CCCAGCGAAAGCGCCCAACACCCCT...GACATCTC...GAGATCGAT 679
233 yProSerGlnLysProProThrProLeuAspIleLeuGlyAspSer 250
680 TCAAAAAGGCGCTCCCTGAG...GTGACCAAGCTCAAGATGACAC 726
250 heLysGlyAlaProValGlnGlyLysProThrPheLysAspGly 266
727 ACCGACGACGCTCTGAGCTTTCATGACATGAGGAGGCGGG 776
267 ThrHisGlnThrPheLeuGlnLeuPheMetTyrLeuAsnGlnVal 283
777 CAGAGTCGCGCTGGCGCTAT...GACATGTCGAGAACGCTTCATG 823
283 yLysHisGlyAlaGlyAlaValAlaPheAspIleValGlnAsnAr 300
824 GAATGAGTCC...CGAGTATTCAC...GAGACCCGACGAGC...ACC 864
300 yAsn***SerProArgGlyIleLeuAlaGlyAlaProAlaGlnHis 316
865 ATCCTTACATGCTCATTTAGACATGAGGCGCTTCAC...ATGAGCG 911
317 SerProTyrHisAlaHisLeuAspIleArgIleuSerProTrpAsp 333
912 G...GAAGTCGGCAAAATCAACAGCGCTGGCTG...AAATTGCTG 955
333 gGlySerGlyLysIleLysGlnGlyProGlyAlaTyrLysPheAla 350
956 AGCTGCTATACCGCTT...CGGCGTAGCGCTGAGTGTGATTTGTC 1002
350 IuLeuValTyrThrGlyPheProAlaGlnAlaProGlnLysGluPhe 365
1003 CGCCACATGC...ATCGCCAAATCCGACGAGCGAGTGGAGG 1040
366 .....CysProProThrGlyIleAlaLysValProArgLysPro 380
1041 GAAAGTCAGAGTCCGCTCCCTCAAGGC...CAGGCTTACA 1078
380 yLysGlyLysValGlnValAlaLysSerPheLysGlyProGlnVal 397
1079 TCCCTC...GGCGGAGTCCCA...CTGCTCTCTACAAATGAGAG 1119
397 IeLeuSerGlyArgGlnIleProHisLysLeuLeuThrMetArgI 413
1120 CTGTGAGCATGAGCTG...CAGGCTGATTT...GAGCCAACTGATG 1163
414 IeValLysHisGlyThrCysLysGlyAspTyrTrpArgProIleAs 430
1164 CACCGGTCATCAACATCAATTCCTCAGGCTGAG...GAAT 1204
430 yHisArgGlySerLysHisGlnIleProPheLysAlaGlnLysNT 447
1205 ATCATCGTCTC...CAGAGCAAGCTACTGCCAA 1236
447 yHisArgLeuProArgAlaLysValThrCysGln 458

seq_name: /cgn2_6/plodata/2/paa/US60_COMB.pep:US-60-185-362-544
seq_documentation_block:
: Sequence 544, Application US/60185362
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Viven
: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS.
: TITLE OF INVENTION: AND USES THEREOF

```

```

: FILE REFERENCE: C1000285
: CURRENT APPLICATION NUMBER: US/60/185,362
: CURRENT FILING DATE: 2000-02-28
: NUMBER OF SEQ ID NOS: 836
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 544
: LENGTH: 299
: TYPE: PRT
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(299)
: OTHER INFORMATION: Xaa = Any Amino Acid
US-60-185-362-544

alignment_scores:
Quality: 1207.50 Length: 300
Percent Similarity: 91.33 Gaps: 1
Percent Identity: 77.667

alignment_block:
US-09-775-693-1 x US-60-185-362-544

Align seg 1/1 to: US-60-185-362-544 from: 1 to: 299

316 GCCACGCGGAGGGGCGCAAGTATGTGTCACGCGCCACAGAGAGG 365
1 AlacInItrGlnGlyAlaLysTyrValSerHisSerAlaMetClyLysol 17
366 GAGCATGAGTCCGGCTTTAGCTCAGCTGCTACTGAGCCGCCAG 415
17 yAsnAspGlnValTyrPheGlnLeuAlaCysTyrSerLeuAlaProGlnI 34
416 TAAAGTCATTCCTCCCTGGAGGATGCTGAATTCATCAACGCTTCA 465
34 IeLysValIleAlaProGlyArgIleProGlnPheTyrLysMetLys 50
466 GCGCGCAATGACTGATGAGTACGCAAGCAACAGCGGATCCATCC 515
51 GlytSerAspLeuMetGlnTyrAlaGlnLysHisGlyIleProIlePr 67
516 GCTACATCCCAAGAACCGTGGAGCATGGATGAGAACCTCATGACATCA 565
67 oValThrLeuLysHisProTrpAspMetAspGlnMetHisIle 84
566 GCTACAGAGCTGGAATCTGAGAACCCCAAGAACGAGCTTCAGGT 615
84 eHisGlnAlaGlyIleLeuGlnAsnProLysAsnGlnAlaProSerGly 100
616 CTCTACACGAGAACCGGACCGCCCAAGCGCCCAACACCTGACAT 665
101 IeTyrMetLysIleGlnAspLeuAlaLysAlaProAsnThrProAsnI 117
666 TCTGAGATGAGTTCAAAAAAGGGGCTCCGTGAAGTGCACAAAGTCA 715
117 ePheLysThrGln...LysLysGlyValProValLysValThrSerIle 133
716 AGGATGACACGACCGACGACCTCTCTGAGCTTTCATGATCACTGAAC 765
133 yAspGlyThrThrHisGlnThrSerLeuValLeuPheMetTyrLeuAsn 149
766 GAGTCGCGGCGCAGACATGGCGTGGCGGCGATTCATGCTGGAGAAC 815
150 GluValAlaGlyLysHisSerValGlnHisIleAspIleValGlnAsn 166
816 CTTCATTGGAATGAGTCCGAGGTATCTACGAGACCCAGCAGGACCA 865
166 gPheIleGlnMetLysSer***GlyIleCysLysThrProAlaGlyThrI 183
866 TCCCTTACCATGCTCATTTAGACATGAGCGCTTCACATGAGCGGGA 915
183 IeLeuTyrHisProHisLeuAspIleGlnIlePheAlaMetGlnGln 199

```

```
916 GTGGCAAAATCAACAAGCCCTGGCTTGAATTGCTGAGCTGTGTA 965
|||||
200 ValArgLysIleLysGlnGlyLeuGlyPheAlaGlnLeuValTy 216
966 TACGGGTTAGGGCCTAGCCCTGAGTGTGATTTGCCGCCACATGCATCG 1015
|||||
216 rThrgLypHetrPhisAsnProGlnCysAspPheAlaHisIscysIleA 233
1016 CCAAGTCCCAAGAGCAGAGTGGAGGAAAGTGCAGTGTCCGTCTCAG 1065
|||||
233 lAlysSerGlnAspArgValGlnGlyValGlnValSerIlePheLys 249
1066 GGCCAGGTGTACATCTCTGGCCGGAGTCCCACTGTCTCTTACAATGA 1115
|||||
250 GlnGlnValTyrlleLeucysGlnGlnProProLeuSerLeuTyrsG 266
1116 GGAGCTGTGTGATGAACGTGCAGGGGTGATTATGAGCCAACTGATGCCA 1165
|||||
266 yGlnGlnValSerMetAsnValGlnGlnAsnAspGlnProValGlyAspT 283
1166 CCGGGTTCATCAACATCAATTCCTCAGGCTGAAGAATAATCATCGTCTC 1215
|||||
283 hrSerLeuIleAsnIleAsnSerLeuArGmetLysGlnTyrlHisHisLeu 299
```



OM of: US-09-775-693-1 to: Pending\_Patents\_AA\_New:\* out\_format: pfs  
Date: Feb 12, 2002 4:07 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame-n2p.model -DEV=rlp  
-Q=/cgn2\_1/USPTO.spool/US0975693/runat\_12022002\_124153\_1396/app\_query.fasta\_1.1315  
-DB=Pending\_Patents\_AA\_New -OPMT=fastan -SUFFIX=rapn  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -RGAPOP=6.000 -RGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=bloms62 -TRANS=human40.coi -LIST=45 -DOCLIGN=200  
-THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US0975693\_@CGN1\_1\_44 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPHY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-775-693-1  
Query length: 1239  
Database: Pending\_Patents\_AA\_New:\*  
Database sequences: 216105  
Search time (sec): 443.280000

## Score list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-981-353-87 + 2123.00	3133.67	2.1e-167	41		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-340-187-701 + 2036.00	3000.91	3.6e-160	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-340-187-701 + 1603.50	2362.27	1.8e-124	5		
/cgn2_6/ptodata/2/paa/US10_NEW.COMB.pcp:US-10-015-127-9692 + 944.00	1384.36	5.8e-70	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-79357 + 866.50	1268.85	1.5e-63	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-79357 + 866.50	1268.34	1.5e-63	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55742 + 845.50	1237.67	8.2e-62	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55742 + 845.50	1237.16	8.2e-62	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-897-516-5764 + 815.00	1199.49	1.1e-59	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-11258 + 815.00	1190.76	2.9e-59	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-675-784-12974 + 759.00	1114.64	9.6e-55	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-79359 + 655.50	959.15	3.7e-46	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55744 + 639.50	935.38	7.7e-45	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-11255 + 598.00	874.72	2.0e-41	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-10-023-386-32991 + 477.00	702.75	1.6e-31	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78396 + 167.50	236.18	7.6e-06	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-53190 + 160.50	225.79	2.9e-05	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-5221 + 156.50	215.42	7.2e-05	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-5220 + 156.50	213.65	7.6e-05	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-5219 + 156.50	213.38	7.7e-05	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-49072 + 153.00	212.97	0.0001	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55640 + 151.50	213.61	0.0002	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78930 + 147.50	207.61	0.0003	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78870 + 147.50	205.34	0.0004	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-52153 + 147.00	202.90	0.0004	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-53785 + 147.00	202.55	0.0004	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-53785 + 146.00	199.94	0.0005	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-10-021-698-145 + 146.00	195.08	0.0006	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78956 + 145.00	203.34	0.0005	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78546 + 145.00	198.46	0.0006	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55291 + 144.50	203.05	0.0006	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55765 + 144.00	201.85	0.0007	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-78381 + 144.00	196.29	0.0008	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-53616 + 140.50	191.10	0.0015	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-70031 + 140.00	197.25	0.0013	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-70031 + 140.00	197.07	0.0014	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-70029 + 140.00	196.98	0.0014	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-55629 + 138.00	192.48	0.0021	5		
/cgn2_6/ptodata/2/paa/US09_NEW.COMB.pcp:US-09-708-427-69207 + 136.50	190.97	0.0027	5		

/cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-708-427-53125 + 136.50 183.93 0.0034  
/cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-708-427-54261 + 135.50 184.27 0.0038  
/cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-708-427-70389 + 134.00 184.27 0.0048  
/cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-708-427-50632 + 133.00 185.41 0.0054  
/cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-708-427-77679 + 132.50 187.74 0.0053

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pcp:US-09-981-353-87

## seq\_documentation block:

; Sequence 87, Application US/09981353  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; TITLE OF INVENTION: GENS EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US  
; CURRENT APPLICATION NUMBER: US/09/981,353  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 87  
; LENGTH: 412  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Incyte ID No: 1543330CD1  
US-09-981-353-87

alignment\_scores:  
Quality: 2123.00 Length: 412  
Ratio: 5.178 Gaps: 0  
Percent similarity: 99.515 Percent identity: 99.272

## alignment block:

US-09-775-693-1 x US-09-981-353-87

Align seg 1/1 to: US-09-981-353-87 from: 1 to: 412

1 ATGTGACCAAGAGCTCCGTTGCTTACAGTGGCGCTGACAC 50  
|||||  
1 MetSerSerIysGlySerValValLeuAlaIyrSerIyGlyLeuAspTh 17  
51 CTGTGACCTCTGCTGTGGCTGAGGACAAAGCTATACCTATGCT 100  
|||||  
17 rSerCysIleLeuValIyrLeuIySgIInGlyIyrAspValIleAlat 34  
101 ATCTGGCCAACTGGCCAGGAAGAGCTCGAGAGCCAGAGAG 150  
|||||  
34 yLeuAlaIasIleGlyIInIySgIInAspPheGlyIInAlaIyIyS 50  
151 GCACTGAAGCTTGGGCGCAAAAGGCTTCATTCAGATGTCAGAGGA 200  
|||||  
51 AlaLeuIySgIInIyAlaIySgIInValPheIleGlyIInAspValSerArgI 67  
201 GTTGTGAGAGAGTTCATTCGCGCCATCCAGTCCAGGAGCTATG 250  
|||||  
67 yPheValIyGlyIInIyPheIleIyrPheIleIleGlyIInSerIleAlaIyG 84  
251 AGGACCGCTACCTCGGCGACCTCTTCGAGGCGCTCGATCCGCC 300  
|||||  
84 IAspArgIyIyLeuLeuGlyIInIyrSerIleAlaIyPheIleAlaIy 100  
301 AAACAAAGTGAATGCCCGCCAGGAGGAGGCGCCAGATGTGTCCAGG 350  
|||||  
101 LySgIInValIleAlaIyIInIyGlyIInIyAlaIyIyValSerIleG 117  
351 CGCCACAGGAAGGAGAGAGTACAGTCCGCTTACGCTACCTACT 400  
|||||  
117 yAlaIyrIyGlyIySgIyAspIyGlyIInValIyPheIyIleuSerIyIyS 134  
401 CACTGGCCCCAGATAAAGGCTATTCCTCGAGAGTGGCTGAATTC 450  
|||||

```

134 erleuAlaProGlnIleLysValIleAlaProTrpArgMetProGluPhe 150
451 TACAACGGTTCAAGGGCCCAATGACCTGATGGATGACCAAGACACA 500
151 TyrSnrMrpPheLysGlyArgAsnAspMetGluTyrAlaLysGlnI 167
501 CGGATTCACATCCCGGTCACTCCCAAGAACCCGTGGACGATGATAGA 550
167 sGlyIleProIleProValThrProLysAsnProTrpSerMetAspGlu 184
551 ACCATGACGACATGACGTACGAGGCTGGATCTCTGGAGAACCCCAAGAC 600
184 snLeuMetHisIleSerTyrGluAlaGlyIleLeuGluAsnProLysAsn 200
601 CAAGCGCTCCAGGTCTCTACAGAGAGACCCAGACCCCAAGCCGCC 650
201 GluAlaProProGlyLeuTyrTrpLysThrGlnAspProAlaLysAlaP 217
651 CAACACCCCTGACATTCGAGATCGAGTTCAAAAAAGGGTCCCTGTGA 700
217 oAsnThrProAspIleLeuGlnIleGluPheLysGlyValProValI 234
701 AGGAGCAACAGTCAGAGATGACACACCCACGACCTCTGAGAGCTC 750
234 yValThrAsnValLysAspGlyThrThiSgInThrSerLeuGluLeu 250
751 TTCATGTACCTGAAGAGTGGCGGCAAGCATGCGGTGGCGGTATGGA 800
251 PheMetTyrLeuAsnGlnIuValAlaGlyLysHisGlyValGlyArgIle 267
801 CATCCGTGGAGAACCGCTTCTGATTTGAATGAAGTCCGAGTATCTAGA 850
267 pIleValGluAsnArgPheIleGlyMetLysSerArgGlyIleTyrGlu 284
851 CCCACAGACGACCATCTTACATGCTCATTTAGACATGAGGCTTC 900
284 hrProAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGluAlaPhe 300
901 ACCATGGACCGGAGAGTGGCAAAATCAACAGGCTGGCTGGAAT 950
301 ThrMetAspArgGluValAlaArgLysIleLysGlnGlyLeuGlysh 317
951 TGCTGAGCGGTGATACGGGTTACGGGCTACCCCTGAGTGTGAATTTG 1000
317 eAlaGluLeuValTyrThrGlyPheTrpHisSerProGluLysGluPhe 334
1001 TCCGCCATGTCATCCGCCAAGTCCCAAGAGCGAGTGAAGGAATGCGAG 1050
334 aAlaArgHisCysIleAlaLysSerGlnGluArgValGlnGlyLysValGln 350
1051 GTGTCCGTCTCAAGGGCCAGGTGATACATCTCGCGCGGAGATCCCACT 1100
351 ValSerValLeuLysGlnValTyrIleLeuGlyArgGlnSerProIle 367
1101 GTCTCTCATGATGAGAGCTGTGAGCATGAAGTGAAGGAGGTGATTTG 1150
367 uSerLeuTyrAsnGlnGluLeuValSerMetAsnValGlnGlyAspTyrG 384
1151 AGCCAACTGATGCCAACCGGGTTTCATCAACATCAATTCCTCAGGCTGAAG 1200
384 LuProThrAspAlaThrGlyPheIleAsnIleAsnSerLeuArgLeuLys 400
1201 GAATATCATCTCTCCAGAGCAAGTCACTGCCAAA 1236
401 GluTyrHisArgLeuGlnSerLysValThrAlaLys 412

```

```

seq_name: /cgn2_6/ptdata/2/paa/us60_NEW_COMB.pep:us-60-340-187-701

```

```

seq_documentation_block:

```

```

; Sequence 701, Application US/60340187
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Zhiwei

```

```

; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Meng, Gezhi
; APPLICANT: Boyle, Bryan J
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 813
; CURRENT APPLICATION NUMBER: US/60/340,187
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/04941
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/US01/08656
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/14827
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 1192
; SOFTWARE: PL_FL_genes Version 6.0
; SEQ ID NO 701
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-340-187-701

```

```

alignment_scores:
  Quality: 2036.00      Length: 397
  Ratio: 5.168          Gaps: 0
  Percent Similarity: 99.244  Percent Identity: 98.741

```

```

alignment_block:

```

```

US-09-775-693-1 x US-60-340-187-701 ..

```

```

Align seg 1/1 to: US-60-340-187-701 from: 1 to: 583

```



```

1  ATGTCCAGCAAGGCTCCGTGCTTCTGAGCTACAGTGGCGGCTGGACAC 50
   |||||||
77  MetSerSerLysGlySerValValLeuAlaIylSerGlyGlyLeuSprn 93
   |||||||
51  CTCGGGATCTCTCGTGGGCGAGCAAGGCTATGACGTGATTCCT 100
   |||||||
93  rSerGlyLeuValIylrLeuLysGlyGlnGlyTyrAspValIleAlar 110
   |||||||
101  ATCTGGCCACATTCGGCCAGAGAGACCTTCAGAGAACCCAGAGAGAG 150
   |||||||
110  yLeuAlaAsnIleGlyGlnLysGlnAspPheGlnGlnAlaArgLys 126
   |||||||
151  GCAGTGAAGCTTGGGGCAAAAGGCTTCATTGAGAGATGTCACACAGGA 200
   |||||||
127  AlaleuLysLeuGlyAlaLysLysValPheIleGlnAspValSerArg 143
   |||||||
201  GTTTTGGGAGGAGTTCATCTGGCCGCGCATCCAGTCCAGCGCAGCTGAT 250
   |||||||
143  upheValGlnGlnupheIleTrrProAlaIleGlnSerSerAlaLeuTyr 160
   |||||||
251  AGGACCGCTACCTCCTGGGACCTCTCTTCAGGCGCTGCATGGCCGC 300
   |||||||
160  LuAsprArgIyrLeuLeuGlyThrSerLeuAlaArgProCysIleAlaArg 176
   |||||||
301  AAGCAGTGAATCGCCCGACGGGGCCAGTATGTGTCCACAGG 350
   |||||||
177  LysGlnValGlnIleAlaGlnArgGlnGlyAlaLysTyrValSerHisG 193
   |||||||
351  CGCCACAGAGAAAGGAGACATCAGGTCGGTTTGAAGCTCAGCTGCTACT 400
   |||||||
193  yAlaThrGlyLysGlyAsnAspGlnValaArgPheGlnLeuSerCysTyr 210
   |||||||
401  CACTGGCCCCAGATAAGGTCATGCTCCTCGAGAGATGGCTGAATTC 450
   |||||||
210  erLeuAlaProGlnIleLysValIleAlaProTrrArgMetProGlnup 226
   |||||||
451  TACACCGGTTCAAGGGCCGCAATGACCTGATGAGATACGCAAGACACA 500
   |||||||
227  TyrAsnArgPheLysGlyArgAsnAspLeuMetGlyTyrAlaLysGlnH 243
   |||||||
501  CGGATTCCTCATCCCGGCTACTCCCAAGAACCCGTGAGCAGTGAATGAGA 550
   |||||||
243  sGlyIleProIleProValThrProLysAsnProTrrSerMetAspGlnA 260
   |||||||
551  ACCTCAGTCACATCAGCTACGAGGCTGGAATCCTGGAGAACCCCAAGAAC 600
   |||||||
260  snLeuMetHisIleSerTyrGlnAlaGlyIleLeuGlnAsnProLysAsn 276
   |||||||
601  CAAGCGGCTCCAGGTCCTACACAGAGACCCAGAGCCAGCCAAAGCCCC 650
   |||||||
277  GlnAlaProProGlyLeuTyrThrLysThrGlnAspProAlaLysAlaPr 293
   |||||||
651  CAACACCCCTGACATTCGAGATCGAGTTCAAAAAGGGGCTCCCTGTGA 700
   |||||||
293  GAsnThrProAspIleLeuGlnIleGlnPheLysGlyValaLysProVal 310
   |||||||
701  AGGTGACCAAGCTGAGGATGGACACCCAGACCTCCTTGGAGAGTC 750
   |||||||
310  yValThrAsnValLysAspGlyThrThrHisGlnThrSerLeuGlnLeu 326
   |||||||
751  TTCATGTACTGAAGAGAGTGGCGGCAAGCATGGCTGGGCGCTGATTGA 800
   |||||||
327  PheMetCylrLeuAsnGlnValaIleGlyHisGlyValaIleArgIleAs 343
   |||||||
801  CATCGTGGAGAACCGCTTCATTTGAATGAATGCCAGGTTTCACAGAGA 850
   |||||||
343  rIleValaGlnAsnArgPheIleGlyMetLysSerArgLysIleIyrGlnT 360
   |||||||
851  CCCGAGAGGACACATCCTTACATGCTCATTTAGACATGAGAGCCCTTC 900
   |||||||
360  hPrrAlaGlyThrIleLeuTyrHisAlaHisLeuAspIleGlnAlaPhe 376
   |||||||
901  ACCATGAGCCGGGAAGTGGCAAAATCAAAACAGGCTGGGCTTGAATTT 950

```

```

|||||
377  ThrMetAspArgGlnValaArgLysIleLysGlnGlyLeuGlyLeuYspH 393
|||||
951  TCGTAGAGTGTATACCGGTTTACGGCCTAGCCCTGAGTGTGAATTTG 1000
|||||
393  eAlaGlnLeuValIylThrGlyPheThrPrrHisSerPrrGlnCysGlnPheV 410
|||||
1001  TCCGCGCTGCATGCCCAAGTCCAGAGAGCGATGAGAGGAAAGTGCAG 1050
|||||
410  AlArgHisCysIleAlaLysSerGlnGlnArgValaGlnGlyLysValGln 426
|||||
1051  GTGTCCGTCCTCAGAGCGCAGGTGATACCTCGGCGGAGTCCACAT 1100
|||||
427  ValSerValLeuLysGlyGlnValTyrIleLeuGlyArgGlnSerProLe 443
|||||
1101  GTCCTCTACATGAGGAGCTGTGAGCATGACATGACGTCAGGCTGATTTAT 1150
|||||
443  uSerLeuTyrAsnGlnGlnLeuValSerMetAsnValaGlnGlyAspTyrG 460
|||||
1151  AGCCAACTGATGCCACCGGTTTCATCAACATTCATTCCTC 1191
|||||
460  LuProThrAspAlaThrGlyPheIleAsnIleLysPheLeu 473
|||||

seq_name: /cgn2_6/ptodata/2/paa/us60_NEW_COMB pep:us-60-340-187-1176

seq_documentation_block:
: Sequence 1176, Application US/60340187
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Weng, Gezhil
: APPLICANT: Boyle, Bryan J
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: TITLE OF INVENTION: Polypeptides
: FILE REFERENCE: 813
: CURRENT APPLICATION NUMBER: US/60/340,187
: CURRENT FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631

```

```

; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/770,160
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/US01/08656
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/14827
; NUMBER OF SEQ ID NOS: 1192
; SOFTWARE: PCT-Genes Version 6.0
; SEQ ID NO 1176
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(433)
; OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set for
; US-60-340-187-1176

```

```

alignment_scores:
    Quality: 1603.50      Length: 428
    Ratio: 4.310          Gaps: 15
    Percent Similarity: 86.916    Percent Identity: 80.140

```

alignment\_block:

US-09-775-693-1 x US-60-340-187-1176 ..

Align seg 1/1 to: US-60-340-187-1176 from: 1 to: 433

```

1 ATGTCACAGCAAGGCTCCGTTCTGCGCTACAGTGGCGGCTGAGAC 50
1 MetSerSerLysGlySerValIValLeuAlaIerIerLysGlyLeuAspH 17
51 CTCGTCATCCTCGTGGCTGAGCAAGCAAGCTATGACGTATGCTCT 100
17 rSerCysIleLeuValIerIerLysGlyLeuIleuIleuIleuIleu 34
101 ATCTGGCCACATTCGCCAGAGCAAGCAAGCTTCGAGAGCCAGAGAG 150
34 yLeuAlaAsnIleGlyGlnLysGlnAspPheGluAlaIleuLysLys 50
151 GCACGTGAAGCTTGGGGCCAAAGGTTGATTGAGAGTGCAGCAGCA 200
51 AlaLeuLysLeuGlyAlaLysLysValPheIleGluAspValSerArg 67
201 GTTGTGGAGAGTTCATCTGCGCGGCAATCCAGTCCAGCGCCTATG 250
67 uPheValIleuGluPheIleuIleuIleuIleuIleuIleuIleuIleu 84
251 AGGACCCCTACCTCTGGGCACTCTTGCAGGCGCTGATCGCCGCG 300
84 LuAspArgIleuLeuGlyThrPhePheAlaArgProCysIleAlaArg 100
301 AACCACTGGAATCGCCGCGGAGGAGGCGCAATATGATGCCAGCG 350
101 LysGlnValIleuIleuAlaGlnArgGluGlyAlaLysLysValSerHis 117
351 CGCCACAGAAAGGAGCAATCAGTCCGTTGAGTTCAGCTCTCTCT 400
117 yAlaThrGlyLysGlyAsnAspGlnValArgPheGluLeuSerCysLys 134
401 CACTGGCCCCCAGATTAAGGTATGCTCCCTGAGAGATGCTGAATTC 450
134 etLeuAlaProGlnIleuLysValIleuAlaProIleuArgMetProGluPhe 150
451 TACACCGGTTCAAGGCGCAATGAGTGAAGTGAAGTGAAGTGAAGTGA 500
151 TyrAsnArgPheLysGlyArgAsnAspLeuMetGluTyrAlaLysGlnHis 167

```

```

501 CGGATTCACATCCGCTGCTACATCCCAAGAACCCGTTGAGCATGATGAGA 550
167 sGlyIleProIleProValThrProLysAsnProIlePheSerMetAspGln 184
551 ACCTGATGACATCAGCTACGAGCGGCAATCTGAGAACCCCAAGAAC 600
184 snLeuMetHisIleSerTyrGluAlaGlyIleLeuGlnAsnProLysAsn 200
601 CAAGCGCTCCAGGCTCTACACAGAACCCAGAGCCAGCCCAAGCCCC 650
201 GlnAlaProProGlyLeuIleuIleuIleuIleuIleuIleuIleuIleu 217
651 CAACACCCCTGACATTCCTGAGATGAGTCAAAAAGGGTCCCTGGA 700
217 oAsnThrProAspIleLeuGlnIleGluPheLysGlyValProValG 234
701 AG...GTGACCAAGTCAAGATGCGCACCCACCCAGACCTCTTGAG 747
234 IuGlyGlyProThrPheLysAspGlyThrThrHisGlnThrPheLeuGln 250
748 CTCTTCATGATACCTGAGCAAGTGGCGGCAAGATGCGGCGGCG...CG 794
251 LeuPheMetTyrLeuAsnGlnValAlaIleLysHisIleValGlyProTy 267
795 TATTGACATCGTGAGAACCGCTTC...ATTGAATGATGCCGAGGTA 841
267 rLeuThrSerTyrGlnAsnArgPheHisIleProGluLeuLysSerArgGly 284
842 TCTACGAGACCCCA...GCAGGCACATCCCTTACATGCTCATTTAGAC 888
284 IeLeuArgArgProGlnAlaIleLysIleuIleuIleuIleuIleuIleu 300
889 ATCGAGGCTTCACCATGAGACCCGGAAGTCCGCAAAATC...AAACAGG 935
301 IleGlnAlaPheThrMetGlyGlyAspArgAlaGlnIleProAsnGlnIle 317
936 CCTGGGCTTGAATTCCTGAGCTG...GTGTTACCGGTTTACGGCTTA 982
317 yLeuGlyPheIleuPheValGlnLeuGlyValTyrArgPheSerGlyThr 334
983 GCCCTGAGTGAATTTGTCGCCGCACTGCATGCCCAAGTCC...CAGGAG 1029
334 IAProGluCysGlnLeuValGlyProCysLeuArgIleSerProGlnIle 350
1030 CGACTGGAAGGAAA...GTGCAGGTTCCTCGTTCAGAGGC...CAGGT 1073
351 ArgValGlnIleLysSerAlaGlyValProSerLeuLysGlyProArgCys 367
1074 GTACATCCTCGCGCGGAGTCCCA...CTGCTCTCTCATGAGAGAC 1120
367 sThrSerLeuGlyProGluValProHisCysSerLeuIleuGlnIle 384
1121 TGSTGAGCATGAAGCTG...CAGGGTGAATTAAGCAACTGAT...GCC 1164
384 euValLysHisGlyThrCysGlnIleAspTyrGluProAsn***LeuPro 400
1165 ACCGGGTATCATCAATCATTCATCCCTCAGCGTGAAG...GAATA 1205
401 ProGlyPheIleGln...ThrSerIleSerLeuLysAlaGlnIleAsnTy 416
1206 TCATCGTCTC...CAGAGCAAGTCACTGCGCAA 1236
416 rHisArgLeuProArgAlaLysValThrAlaGln 427

```

seq\_name: /cgn2\_6/ptdata/2/paa/US10\_NEW\_COMB pep-US-10-015-127-9692

seq\_documentation\_block:

```

; Sequence 9692, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

```



LOCATION: 1..425  
OTHER INFORMATION: Ceres Seq. ID 1965189  
US-09-708-427-79358

alignment\_scores:  
Quality: 866.50 Length: 393  
Ratio: 2.947 Gaps: 5  
Percent Similarity: 74.809 Percent Identity: 45.547

alignment\_block:  
US-09-775-693-1 x US-09-708-427-79358 ..

Align seg 1/1 to: US-09-708-427-79358 from: 1 to: 425

```

19 GTGGTCTGCTTACAGTGGCGGCTTGAGACACCTGTCATCCCTGCTG 68
   |||||.....|.....|.....|.....|.....|.....|.....|
30 ValValLeuAlaIleTyrSerGlyLeuAspThrSerValIleValProtr 46
   |||||.....|.....|.....|.....|.....|.....|.....|
69 GCGAAGAGAACAA...GGCTATGACGTATTCCTATCTGGCCAACTTG 115
   |||||.....|.....|.....|.....|.....|.....|.....|
46 pleuAtgGluAsnTyrGlyCysGluValAlaCysPheThrAlaAspValG 63
   |||||.....|.....|.....|.....|.....|.....|.....|
116 GCCAGAGAA...GACTTCGAGAGAACGACGAGGAGGAGTTTGAGAGA 162
   |||||.....|.....|.....|.....|.....|.....|.....|
63 LylGlnGlyAlaIleGluLeuGlnGlyLeuGlnLysAlaLysAlaSer 79
   |||||.....|.....|.....|.....|.....|.....|.....|
163 GGGGCCAAAGAGGTTTCATGAGAGATGTCAGAGGAGGTTTGAGAGA 212
   |||||.....|.....|.....|.....|.....|.....|.....|
80 GlyAlaCysGlnLeuValAlaLysAspLeuLysGlnGlnLysAlaSer 96
   |||||.....|.....|.....|.....|.....|.....|.....|
213 GTTCATCTGGCGGCGCATCCAGTCGACGACGACGTATGAGAGACCGTACC 262
   |||||.....|.....|.....|.....|.....|.....|.....|
96 uTyrIleTyrProCysLeuArgAlaGlyAlaValTyrGlnAlaTyrLys 113
   |||||.....|.....|.....|.....|.....|.....|.....|
263 TCCGCGGCACTCTCTTCCAGGCGCTGTCATCCCGGCAACAGAGGAA 312
   |||||.....|.....|.....|.....|.....|.....|.....|
113 euLeuGlnTyrSerMetAlaArgProValIleAlaLysAlaMetValAsp 129
   |||||.....|.....|.....|.....|.....|.....|.....|
313 ATCCGCCAGGCGGAGGGGCGCAGTATGTGTCCACGCGCCACAGAGAA 362
   |||||.....|.....|.....|.....|.....|.....|.....|
130 ValAlaLysGlnValAlaGlyAlaAspAlaValAlaHisGlyCysThrGly 146
   |||||.....|.....|.....|.....|.....|.....|.....|
363 GGGGAGAGATCAGTCGCGGTTTGGCTGACGTCGTCATCTAGCGGCCCC 412
   |||||.....|.....|.....|.....|.....|.....|.....|
146 sGlyAsnAspGlnValAlaArgPheGluLeuThrPheTyrAlaLeuAsnPro 163
   |||||.....|.....|.....|.....|.....|.....|.....|
413 AGATAAAGTCATGTGCTCCCTGGAGATGCTGAATTCTACAAACGGTTC 462
   |||||.....|.....|.....|.....|.....|.....|.....|
163 IuLeuLysValAlaIleProtrPargLutPasp.....Ile 175
   |||||.....|.....|.....|.....|.....|.....|.....|
463 AAGGCGCGAATGACCTGATGAGTACCAAGCAACAGCGGATTCAT 512
   |||||.....|.....|.....|.....|.....|.....|.....|
176 ThrGlyArgGluAspAlaIleGluTyrAlaLysLysHisAsnValProva 192
   |||||.....|.....|.....|.....|.....|.....|.....|
513 CCCGGTCACTCCCAAGAACCCGTGAGAGATGAGAACTCATGACA 562
   |||||.....|.....|.....|.....|.....|.....|.....|
192 LProValSerLysLysSerIleTyrSerArgAspArgAsnLeuThrPstL 209
   |||||.....|.....|.....|.....|.....|.....|.....|
563 TCAGCTAGAGGCTGATCCTGGAGAACCCCAAGAACCAAGCCCTCCA 612
   |||||.....|.....|.....|.....|.....|.....|.....|
209 euSerHisGlnGlyLysPheLeuGlnAspProAlaAsnGluProLysGlu 225
   |||||.....|.....|.....|.....|.....|.....|.....|
613 GGTCTCTACAGAACCCAGACCCAGCCAAAGCCCGCCCAACCCCTGA 662
   |||||.....|.....|.....|.....|.....|.....|.....|
226 AspMetTyrMetSerIleAlaProGluAsnAlaProSerLysProGlu 242
   |||||.....|.....|.....|.....|.....|.....|.....|
663 CATTCCTCAGATCGATTCAAAAAGGGTCCCTGTGAAGGTGACCAAG 712
   |||||.....|.....|.....|.....|.....|.....|.....|
242 uTyrLeuGlnIleGlyIleIleAlaGlyValProValSerIle...Asn 258
   |||||.....|.....|.....|.....|.....|.....|.....|
713 TCAAGATGCGACCCAGCCACTCTTGAGGCTTCATGATGATCG 762
   |||||.....|.....|.....|.....|.....|.....|.....|

```

```

258 LyrArgAsp.....LeuSerProAlaSerLeuAlaGluLeu 270
763 AAGGAAGTCGCGGCGCAAGCATGCGCTGGCGCTTATGACATGTCGAGAA 812
   |||||.....|.....|.....|.....|.....|.....|.....|
271 AsnGluIleGlyGlyLysHisGlyIleGlyArgIleAspMetValGluAs 287
   |||||.....|.....|.....|.....|.....|.....|.....|
813 CCGCTCATTTGAGATGAGTCCCGAGGTATCTACAGACAGCCAGCAGCA 862
   |||||.....|.....|.....|.....|.....|.....|.....|
287 nArgLeuValCysMetLysSerArgIleValTyrGlnThrProGlyLys 304
   |||||.....|.....|.....|.....|.....|.....|.....|
863 CCATCCTTTACCATGCTCATTTAGACATGAGGCGCTTCACCATGACCG 912
   |||||.....|.....|.....|.....|.....|.....|.....|
304 hTleMetAlaIleAlaValAlaArgGluLeuGlnSerLeuThrLeuAsp 320
   |||||.....|.....|.....|.....|.....|.....|.....|
913 GAAGTCGCGCAATCAACAAAGCGCTTGAATTTGCTGAGCTGT 962
   |||||.....|.....|.....|.....|.....|.....|.....|
321 GluThrMetGlnThrLysAspIleIleAlaLeuLysTyrAlaGluLeuVa 337
   |||||.....|.....|.....|.....|.....|.....|.....|
963 GTATACCGGTTTACGCGCTTACCGCTGAGTGTGAATTTGTCGCGCATGCA 1012
   |||||.....|.....|.....|.....|.....|.....|.....|
337 LTrpAlaGlyArgTyrPheAspProLeuArgGlnSerPheAspAlaPheM 354
   |||||.....|.....|.....|.....|.....|.....|.....|
1013 TCGCCCAAGTCCCGAGAGAGAGTGAAGAGAAAGTGCAGGTGCTGCTC 1062
   |||||.....|.....|.....|.....|.....|.....|.....|
354 eGluLysValThrAlaThrThrGlySerValThrLeuLysLeuTyr 370
   |||||.....|.....|.....|.....|.....|.....|.....|
1063 AAGGCGCAGGTGATCATCTCGCGCGGAGTCCCGCATGCTCTCTACAA 1112
   |||||.....|.....|.....|.....|.....|.....|.....|
371 LysGlySerValAsnValAlaSerArgLysSerProTyrSerLeuTyr 387
   |||||.....|.....|.....|.....|.....|.....|.....|
1113 TGAGGAGCTGCTGAGCATGACGTGCAAGGCTGATTTGAGCCCACTGATG 1162
   |||||.....|.....|.....|.....|.....|.....|.....|
387 gGluAspLysSerSerPheGlnAsnGlyIuIleTyrAsnGlnAlaAspA 404
   |||||.....|.....|.....|.....|.....|.....|.....|
1163 CCACCGGTTTCATCAACATCATTCCTC 1191
   |||||.....|.....|.....|.....|.....|.....|.....|
404 laGlnGlyPheIleArgLeuTyrGlyLeu 413
   |||||.....|.....|.....|.....|.....|.....|.....|

```

seq\_name: /cgn2.6/plodata/2/paa/us09\_NEW\_COMB.pep:US-09-708-427-79357

```

seq_documentation_block:
; Sequence 79357, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79357
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..446
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc.feature
; LOCATION: 1..446
; OTHER INFORMATION: Ceres Seq. ID 1965188
; US-09-708-427-79357

```

alignment\_scores:

Quality: 866.50 Length: 393  
Ratio: 2.947 Gaps: 5  
Percent Similarity: 74.809 Percent Identity: 45.547

alignment\_block:

US-09-775-693-1 x US-09-708-427-79357 ..



```

163 GGGGCCAAAAGGTGTCATTCAGATGTCAGAGGAGGATTTGTGAGGA 212
||||| : : : : : : : : : : : : : : : : : : : : : : : :
80 GLYAlaCysGlnLeuValValLysAspLeuLysGlnGluPheValSerGI 96
||||| : : : : : : : : : : : : : : : : : : : : : : : :
213 GTTCATCTGGCCGGCCATCCAGTCAGCGCAGCTGTATGAGAACCGCTACC 262
||||| : : : : : : : : : : : : : : : : : : : : : : : :
96 uTYrIleTYrProCysLeuArgIaGlyAlaValTYrGlnArgLysTYrL 113
||||| : : : : : : : : : : : : : : : : : : : : : : : :
263 TCCTGGGACCTCTCTTGGCAGGCGCTGATGCGCCGCAACAGAGGAA 312
||||| : : : : : : : : : : : : : : : : : : : : : : : :
113 euleuGIYThrSerMetAlaArgProValIleAlaLysAlaMetValAsp 129
||||| : : : : : : : : : : : : : : : : : : : : : : : :
313 ATGCCCGAGGGAGGGGCGGCAAGTATGTGTCACGCGCCAGAGAA 362
||||| : : : : : : : : : : : : : : : : : : : : : : : :
130 ValAlaLysGlnValAlaGlyAlaAspAlaAlaHisGlyCysThrLys 146
||||| : : : : : : : : : : : : : : : : : : : : : : : :
363 GGGGAACGATCAGTCAGTCGGTTTGTAGCTCAGCTCAGTCCAGTCCGCC 412
||||| : : : : : : : : : : : : : : : : : : : : : : : :
146 sGIYAsnAspGlnValArgPheGluLeuThrPheTYrAlaLeuAsnProG 163
||||| : : : : : : : : : : : : : : : : : : : : : : : :
413 AGATAAGTCATGTCTCCCTGGAGCATGCGCTGAATTCACAAACCGCTTC 462
||||| : : : : : : : : : : : : : : : : : : : : : : : :
163 IuLeuLysValAlaAlaProTrpArgGluTrpAsp.....Ile 175
||||| : : : : : : : : : : : : : : : : : : : : : : : :
463 AAGGGCGGCAATGACCTGATGAGTACGCAAGCAACACAGGAGTCCCAT 512
||||| : : : : : : : : : : : : : : : : : : : : : : : :
176 ThGIYArgGlnAspAlaIleGluTYrAlaLysLysHisAsnValProVa 192
||||| : : : : : : : : : : : : : : : : : : : : : : : :
513 CCGGTCACATCCCAAGAACCCCGGAGCATGATGAGAACCTTCATCACA 562
||||| : : : : : : : : : : : : : : : : : : : : : : : :
192 IProValSerLysLysSerLetyrSerArgAspArgAsnLeuTrpHisI 209
||||| : : : : : : : : : : : : : : : : : : : : : : : :
563 TCAGCTACAGAGGCTGCAATCTTGAGAACCCCAAGAACAGGCGCTCCA 612
||||| : : : : : : : : : : : : : : : : : : : : : : : :
209 euserHisGlnGlyAspIleLeuGlnAspProAlaAsnGluProLysGln 225
||||| : : : : : : : : : : : : : : : : : : : : : : : :
613 GGTCTCTACAGAAAGCCAGACCCAGCCAGCCCAAGCCCAACCCCTGA 662
||||| : : : : : : : : : : : : : : : : : : : : : : : :
226 AspPheTYrMetSerIleAlaProGlnAsnAlaProSerLysProGI 242
||||| : : : : : : : : : : : : : : : : : : : : : : : :
663 CATCTCTGAGATCGAGTTCAAAAAGGGTCCCTGTGAGAGTACCAACG 712
||||| : : : : : : : : : : : : : : : : : : : : : : : :
242 uTYrLeuGlnIleGlyIleIleAlaGlyAlaProValSerIle...Asn 258
||||| : : : : : : : : : : : : : : : : : : : : : : : :
713 TCAAGATGGCACCCACCACCACTCTTGAGAGCTCTTCATGATACCTG 762
||||| : : : : : : : : : : : : : : : : : : : : : : : :
258 LYArgAsp.....LeuSerProAlaSerLeuLeuAlaGluLeu 270
||||| : : : : : : : : : : : : : : : : : : : : : : : :
763 AAGCAAGTCGGGAGCATGGCGGTGGGCGTATTGACATCGTGAGAA 812
||||| : : : : : : : : : : : : : : : : : : : : : : : :
271 AsnGlnIleGlyLysHisGlyIleGlyArgIleAspMetValLys 287
||||| : : : : : : : : : : : : : : : : : : : : : : : :
813 CCGCTTCATTGGAATGAGTCCGAGGTATCTACGAGACCCAGACGGA 862
||||| : : : : : : : : : : : : : : : : : : : : : : : :
287 nArgLeuVal**MetLysSerArgGlyValTYrGlnThrProGlnGly 304
||||| : : : : : : : : : : : : : : : : : : : : : : : :
863 CCATCTTACCATGCTCATTTAGACATGAGGCTTCACCATGAGACGG 912
||||| : : : : : : : : : : : : : : : : : : : : : : : :
304 hIleMetAlaAlaAlaValArgIleuGlnSerLeuThrLeuAspArg 320
||||| : : : : : : : : : : : : : : : : : : : : : : : :
913 GAAGTCGCAAAATCAACAAGCGCTGGAGCTGAAATTTGCTGAGTGTG 962
||||| : : : : : : : : : : : : : : : : : : : : : : : :
321 GluThrMetGlnTrpLysAspIleIleAlaLeuLysTYrAlaGluLeuVa 337
||||| : : : : : : : : : : : : : : : : : : : : : : : :
963 GTATACCGGTTTACGGCTAGCCCTGAGTGAATTTGCGCCACTGCA 1012
||||| : : : : : : : : : : : : : : : : : : : : : : : :
337 ITYrAlaGlyArgTrpPheAspProLeuArgGlnSerPheAspAlaPhe 354
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1013 TCGCAGAGTCCGAGAGCGAGTGAAGGAAATGCGAGTGTCCGTCTC 1062
||||| : : : : : : : : : : : : : : : : : : : : : : : :
354 etGIuLysValIThrAlaIThrThrLysValThrLeuLysLeuTYr 370

```

```

1063 AAGGCCAGGTGATCATCTGCGCGGAGTCCCACTGTCTCTACAA 1112
||||| : : : : : : : : : : : : : : : : : : : : : : : :
371 LysGIserValAsnValAlaSerArgLysSerProTYrSerLeuTYr 387
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1113 TGAGAGCTGTGATGATGAACTGACAGGCTGATTTAGAGCACTGATG 1162
||||| : : : : : : : : : : : : : : : : : : : : : : : :
387 gGIuAspIleSerSerPheGlnAsnGlyLysIleTYr.AsnGln**Thr 403
||||| : : : : : : : : : : : : : : : : : : : : : : : :
1163 CCACCGGTTTCATCA 1177
||||| : : : : : : : : : : : : : : : : : : : : : : : :
404 ProArgGlySerSer 408

```

```
seq_name: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep:US-09-708-427-55742
```

```
seq_documentation_block:
```

```

; Sequence 55742: Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55742
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc:feature
; LOCATION: 1..447
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc:feature
; LOCATION: 1..447
; OTHER INFORMATION: Ceres Seq. ID 1935961
US-09-708-427-55742

```

```
alignment_scores:
```

```

Quality: 845.50 Length: 389
Ratio: 2.936 Gaps: 6
Percent Similarity: 74.036 Percent Identity: 45.501

```

```
alignment_block:
```

```
US-09-775-693-1 x US-09-708-427-55742 ..
```

```
Align seg 1/1 to: US-09-708-427-55742 from: 1 to: 447
```

```

19 GTGTTTGTGCTACAGTGGCGGCTGGACACTGTCATCTCTGTG 68
||| ||||| : : : : : : : : : : : : : : : : : : : :
51 Val**LeuAlaTYrSerGIgLYLeuAspThrSerValIleValProTr 67
||||| : : : : : : : : : : : : : : : : : : : : : : : :
69 GCTGAAGAACA...GGCTATGACGTGCTATGCTATGTCGCAACATG 115
||||| : : : : : : : : : : : : : : : : : : : : : : : :
67 pLeuArgGlnAsnTYrGIgLYCysGlnValValLysPheThrAlaAsp 84
||||| : : : : : : : : : : : : : : : : : : : : : : : :
116 GCCAAGAAGA...GACTTGAGAGAGCCAGAGAAGACGCTGAGCTT 162
||||| : : : : : : : : : : : : : : : : : : : : : : : :
84 IYgInGIYAlaIleGIuLeuGlnGlyLeuGlnLysValAlaLysAlaSer 100
||||| : : : : : : : : : : : : : : : : : : : : : : : :
163 GGGGCCAAAAGGTGTCATTCAGATGTCAGAGGAGGATTTGTGAGGA 212
||||| : : : : : : : : : : : : : : : : : : : : : : : :
101 GLYAlaCysGlnLeuValValLysAspLeuLysGlnGluPheValSerGI 117
||||| : : : : : : : : : : : : : : : : : : : : : : : :
213 GTTCATCTGGCCGGCCATCCAGTCAGCGCAGCTGTATGAGAACCGCTACC 262
||||| : : : : : : : : : : : : : : : : : : : : : : : :
117 uTYrIleTYrProCysLeuArgIaGlyAlaValTYrGlnArgLysTYrL 134
||||| : : : : : : : : : : : : : : : : : : : : : : : :
263 TCCTGGGACCTCTCTTGGCAGGCGCTGATGCGCCGCAACAGAGGAA 312
||||| : : : : : : : : : : : : : : : : : : : : : : : :
134 euleuGIYThrSerMetAlaArgProValIleAlaLysAlaMetValAsp 150

```

```

313 ATCCGCCAGCGGAGGCGGCAAGTATGTCCTCCAGCGGCGGCAAGGAAA 362
      ::::::::::: ||||| ||::::|||::::|::::|::::| 167
151 VALAIALysGluValGlyAlaAspAlaValAlaHisGlyCysThrGly 167
      |||||::|::|::|::|::|::|::|::|::|::|::|::| 412
363 GGGAGACGATGAGTCCGGTTTGAGCTCAGCTGCTACTGCGCCGCC 412
      |||||::|::|::|::|::|::|::|::|::|::|::|::| 167
167 sGlysnAspGlnValAlaArgHeGluLeuThrPheThrAlaLeuAsnProG 184
      ::::::::::: ||::::|||::::|::::|::::| 413
413 AGATAAGGTCATGCTGCTGCGGAGATGCTGAATTCTACACCGGCTTC 462
      ::::::::::: ||::::|||::::|::::|::::| 184
184 LuleuLysValAlaAlaProTPrArgGluTrpAsp.....lle 196
      ::::::::::: ||::::|||::::|::::|::::| 463
463 AAGGCGCGCAATGACCTGATGAGTACGCAAGCAACGCGATTCCAT 512
      |||||::|::|::|::|::|::|::|::|::|::|::|::| 197
197 ThrGlyArgGluAspAlaIleGluThrAlaLysLysHisAsnValProVa 213
      ::::::::::: ||::::|||::::|::::|::::| 513
513 CCCGCTACTCCCAAGAACCCCTGAGCATGAGTGAAGAACTCATGACA 562
      ::::::::::: ||::::|||::::|::::|::::| 213
213 LProValSerLysLysSerIleTyrSerArgAspArgAsnLeuTrpHisL 230
      ::::::::::: ||::::|||::::|::::|::::| 563
563 TCAGCTACGAGCGTGAATCCTGGAGAACCCCAAGACCAACGCGCTCCA 612
      ::::::::::: ||::::|||::::|::::|::::| 230
230 euserHisGluGlyAspIleLeuGluAspProAlaAsnGluProLysGlu 246
      ::::::::::: ||::::|||::::|::::|::::| 613
613 GGTCTACACGAGACGACGACGCAAGCCCAAGCCCAACGCGCTTGA 662
      ::::::::::: ||::::|||::::|::::|::::| 247
247 AspMetTyrMetMetSerIleAlaProGluAsnAlaProSerLysProL 263
      ::::::::::: ||::::|||::::|::::|::::| 663
663 CATTCGAGATCGAGTCAAAAAGGCTCCCTGTAAGTGAACGACG 712
      ::::::::::: ||::::|||::::|::::|::::| 263
263 uTyrLeuGluIleGlyIleIleAlaGlyAlaProValSerIle...AsnG 279
      ::::::::::: ||::::|||::::|::::|::::| 713
713 TCAGAGATGGCACACCCACCAACACCTCTGGAGCTCTGATGATCTG 762
      ::::::::::: ||::::|||::::|::::|::::| 279
279 LyrArgSp.....LeuSerProAlaSerLeuLeuAlaGluLeu 291
      ::::::::::: ||::::|||::::|::::|::::| 763
763 AAGCAAGTCGCGGCAAGCATGGCTGGCGCTGATTCATGCTGGAGAA 812
      ::::::::::: ||::::|||::::|::::|::::| 292
292 AsnGluIleGlyLysHisGlyIleGlyArgIleAspMetValGluAs 308
      ::::::::::: ||::::|||::::|::::|::::| 813
813 CCGCTTCATTCGATGATGATCCGAGGATCTTCGAGACCCCGACGCA 862
      ::::::::::: ||::::|||::::|::::|::::| 308
308 nArgLeuVal***MetLysSerArgGlyValTyrGluThrProGlyLys 325
      ::::::::::: ||::::|||::::|::::|::::| 863
863 CCATCCCTTACCATGCTCATTTAGACATGAGGCTTCACCATGACGCG 912
      ::::::::::: ||::::|||::::|::::|::::| 325
325 hTrrIleMetAlaAlaAlaValAlaArgIleuLysIleuSerLeuThrLeuAspArg 341
      ::::::::::: ||::::|||::::|::::|::::| 913
913 GAAGTGGCAAAATCAACAAGCGCTGGGCTTGAATTTGCTGAGCTGT 962
      ::::::::::: ||::::|||::::|::::|::::| 342
342 GluThrMetGlnTrpLysAspIleIleAlaLeuLysTyrAlaGluLeuVa 358
      ::::::::::: ||::::|||::::|::::|::::| 963
963 GTATACCGGTTTACGGCTTACGCTGAGTGTGAATTTGCTGCGCACGCA 1012
      ::::::::::: ||::::|||::::|::::|::::| 358
358 lTyrAlaGlyArgTrpPheAspProLeuArgGlnSerPheAspAlaPheM 375
      ::::::::::: ||::::|||::::|::::|::::| 1013
1013 TCGCCAAGTCCCAAGAGGAGTGAAGCAAGTGCAGTGTCCGCTCTC 1062
      ::::::::::: ||::::|||::::|::::|::::| 375
375 etGluLysValThrAlaThrThrThrLysSerValThrLeuLysLeuTyr 391
      ::::::::::: ||::::|||::::|::::|::::| 1063
1063 AAGGCGCAGGTGATCATCTCGCGCGGAGAGTCCCACTGCTCTCTACAA 1112
      ::::::::::: ||::::|||::::|::::|::::| 392
392 LysGlySerValAsnValAlaSerArgLysSerProTyrSerLeuTyrAr 408
      ::::::::::: ||::::|||::::|::::|::::| 1113
1113 TGAGAGAGCTGTGAGCATGAACGTGACGAGGTGATATGAGCGCACTGATG 1162
      ::::::::::: ||::::|||::::|::::|::::| 408
408 gLluAspLysSerPheGluAsnGlyGluIleTyr.AsnGln***Thr 424
      ::::::::::: ||::::|||::::|::::|::::| 1163
1163 CCACGCGGTTCATCA 1177
      ::::::::::: ||::::|||::::|::::|::::| 425
425 ProArgLysSerSer 429

```

seq\_name: /cqn2\_6/ptodata/2/paa/US09\_NEW\_COMB pep: US-09-897-516-5764

```

seq_documentation_block:
; Sequence 5764, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Cordin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spilidonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; PRIORITY FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIORITY FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 6409
; SEQ ID NO 5764
; LENGTH: 410
; TYPE: PR
; ORGANISM: Xenorhabdus sp.
US-09-897-516-5764

```

```

alignment_scores:
  Quality: 819.50      Length: 405
  Ratio: 2.836         Gaps: 6
  Percent Similarity: 71.358      Percent Identity: 45.185

```

alignment\_block:

US-09-775-693-1 x US-09-897-516-5764 ..

Align seg 1/1 to: US-09-897-516-5764 from: 1 to: 410

```

19 GTGCTTGGCGCTACAGTGGGCGCTGAGACACCTGCTGATCCGCTG 68
      ::::::::::: |||||::|::|::|::|::|::|::|::| 15
15 lIleValIleuAlaTyrSerGlyLeuAspTrpSerAlaIleIleProTr 31
      ::::::::::: ||::::|||::::|::::|::::| 69
69 GGTGAAGGAAACA.....GCTATGACGTCATTCCTATCTGGCCANCA 112
      ::::::::::: ||::::|||::::|::::|::::| 31
31 pLeuLysGluHisTyrGlyAsnGlyAspValIleAlaPheValAlaAspV 48
      ::::::::::: ||::::|||::::|::::|::::| 113
113 TTGGCCAG...AAGAGACCTTCGAGAGAACCGAGAAAGACGACTGAG 159
      ::::::::::: ||::::|||::::|::::|::::| 48
48 aGlyGlnSerArgGlnAspLeuGluGlyIleGluGlnLysAlaLeuArg 64
      ::::::::::: ||::::|||::::|::::|::::| 160
160 CTTGGGCGCCAAAAGGTGTCATTCGAGGATGACAGGAGATTGTGGA 209
      ::::::::::: ||::::|||::::|::::|::::| 65
65 SerGlyAlaSerGluCysHisIleAlaAspLeuArgGluGluPheIle 81
      ::::::::::: ||::::|||::::|::::|::::| 210
210 GGAGTCATCTGCGCGCCATCCAGTCCAGCGCACTGATGAGACCGCT 259
      ::::::::::: ||::::|||::::|::::|::::| 81
81 sGluTyrValIyrProValLeuLysThrGlyAlaLeuTyrGluGlySerT 98
      ::::::::::: ||::::|||::::|::::|::::| 260
260 ACCTCTGGCACTCTCTTGGCCAGCGCTGATGCGCGCAAAAGAGTG 309
      ::::::::::: ||::::|||::::|::::|::::| 98
98 yTrLeuLeuGlyThrSerMetAlaArgProIleIleAlaLysAlaGlnVal 114
      ::::::::::: ||::::|||::::|::::|::::| 310
310 GAATGCGCCAGGCGGAGGCGCAAGTATGTCCTCCAGCGCGCACAGG 359
      ::::::::::: ||::::|||::::|::::|::::| 115
115 GluLeuAlaLeuLysValAlaAlaAspAlaLeuAlaHisGlyAlaThrCl 131
      ::::::::::: ||::::|||::::|::::|::::| 360
360 AAGGCGAAGCATCAGTCCGCTTGGAGCTCAGCTGCTACTGCGGCC 409
      ::::::::::: ||::::|||::::|::::|::::| 131
131 yLysGlyAsnAspGlnValAlaArgHeGluSerThrTyrThrAlaLeuAlaP 148
      ::::::::::: ||::::|||::::|::::|::::| 410
410 CCGAGATAAAGGTGATTCCTGCTGAGAGATGCTGAATTTACAAACGG 459
      ::::::::::: ||::::|||::::|::::|::::| 148
148 roHisLeuLysValAlaAlaProTPrArgGluTrpAsp..... 160
      ::::::::::: ||::::|||::::|::::|::::| 460
460 TTCAGGCGCGCAATGACCTGATGAGTACGCAAAACACGCGGATTC 509

```



```

161 LeuArgSerArgLalLeuLeuAspTyrLeuLysValAlaArgSpLier 177
510 CATCCCGGTCACCTCCAGAACCCCGTGAGCATGATGAGAACCTTCATGC 559
177 othrThrAlaThrLeuGluLysLleTyrSerArgAspGluAsnAlaTrpH 194
560 ACATGACGCTACGAGGCTGAGATCCCTGAGAACCCCAAGAACGAGCCT 609
194 lslleSerThrGluGlyValLeuGluAsnThrTrpAsnAlaAlaAsn 210
610 CCAAGCTCTACACGAGAACCCAGAGCCCAAGCCCAAGCCCAAGCC 659
211 LysAspCysTrpGluTrpThrValGluProGluAspAlaProAspGluPr 227
660 TGCATTCCTCGAGATCGATTCAAAAAGGGTCCCTGGAAGGTGACCA 709
227 oGluTyrLeuSerValThrValGluLysGluValAlaGlyVal...A 243
710 AGCTCAGAGATGACACACCCACACACCTCCTGAGACCTCTCATGTAC 759
243 sngLysLleLeuSerProTyrGlnCysLeuAsnGlu..... 255
760 CTGAGCAGAGATCGCGGCGACAGCATGGCGGCGCTATTGACATGTGA 809
256 LeuAsnGluLeuGlyAlaLysHisGlyLleGlyArgLleAspLleValGI 272
810 GAACGCTTCATTGGAATGAATGCCGAGTATCTACGAGAACCCACAG 859
272 uAsnArgLeuValGlyMetLysSerArgLysCysTyrGluThrProGlyG 289
860 GCACCATCCTTACCATGCTCATTTAGACATCGAGGCGCTTCACCATG 909
289 lYThrLleMetMetAlaLeuArgGlyLleGluGluLeuValLeuAsp 305
910 CGGAGAGTGGCGAAATCAACAGAGCGCTGGCTTGAATTTGCTGAGCT 959
306 ArgAspSerPheLysTrpArgGluGluLeuGlyLeuGluMetSerTyrVa 322
960 GGTGTATACCGGTTTACGCGCTAGCCCTGAGATGTGATTTGTCGCG 1009
322 lValTyrAspGlyArgTrpPheValProLeuArgGlnSerLleGluAla 339
1010 GCATCGCCAACTCCAGAGAGCGAGTGAAGGAAAGTGCAGGTCCGTC 1059
339 lAlaGluMetLeuAlaGluAsnValSerGlyGluValLleLeuLysLeu 355
1060 CTCAAGGCGAGGTGTACATCTCTGGCGGAGTCCACACTGCTCTCTA 1109
356 TyrLysGlyGluValThrAlaThrGluLysThrSerProHisSerLeuTy 372
1110 CAATGAGAGCTGGTGAAGCATGAAGCTGAGGATGATTATGAGCACA 1159
372 rSerGluGluPheAlaThrPheGlyGluAspGluValTyrAspHisSerH 389
1160 ATGCCACCGGTTTCATCAACATCATTCCTCAGAGCTGAAGAAATATCAT 1209
389 lAlaGluGlyGluValThrAlaThrGluLysThrSerProHisSerLeuTy 404
1210 CGTCTCAGAGCAG 1224
405 AlaleuLysSerLys 409

seq_name: /cgn2_6/ptodata/2/paa/US09_NBM_COMB.pep:us-09-708-427-11258
seq_documentation_block:
; Sequence 11258, Application us/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: us/09/708,427

```

```

; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11258
; LENGTH: 498
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..498
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..498
; OTHER INFORMATION: Ceres Seq. ID 1822815
US-09-708-427-11258

alignment_scores:
Quality: 815.00 Length: 396
Ratio: 2.890 Gaps: 5
Percent Similarity: 71.212 Percent Identity: 42.172

alignment_block:
US-09-775-693-1 x US-09-708-427-11258 ..
Align seg 1/1 to: US-09-708-427-11258 from: 1 to: 498

19 GTGGTCTGTGCTACAGATGGCGGCTGAGACCTGTCATCTCGTGTG 68
128 ValValLeuAlaThrSerGlyGlyLeuAspThrSerValLleValProTr 144
69 GCTGAGGAGCA...GGCTATACGTCATTTGCTTATCTGCCACATTG 115
144 PheLysGluAsnTyrGlyCysGluValAlaCysPheThrAlaAspValG 161
116 GCCAGAG...GAAGCTTCAGAGAACCCAGAGAACGAGCAAGCACTT 162
161 LysGlnGlyLleLysGluLeuGluGlyLeuGluGlnLysAlaLysAlaSer 177
163 GGGGCCAAAAGGTTCATTGAGATGTCAGCAGGAGCTTGTGAGAGA 212
178 GlyAlaSerGlnLeuValAlaLysAspLeuThrGluLysPheValLys 194
213 GTTCATCTGGCGGCTATCCAGTCAGCGCAGCTGATATGAGACCGCTAC 262
194 PheLleIlePheProCysLeuArgAlaGlyAlaLleTyrGluArgLysTyrL 211
263 TCCTGGCACTCTCTTGCAGAGCCCTGCATCGCCGCAACAGTGGAA 312
211 euLeuGlyThrSerMetAlaArgProValLleAlaLys..... 223
313 ATGCCACCGAGGAGGGGCGCAAGTATGTGTCCACGCGGCCACAGAAA 362
223 ..... 223
363 GGGGAGCATCAGGTCCGTTTGAGTCAGCTGTACTACATCGGCCGCC 412
224 .....ValArgPheGluLeuThrPheSerLeuAsnProG 236
413 AGATAAGGTCAATGCTCCCTGAGAGATGCCCTGATTTACAAACGGTTC 462
236 luleLysValAlaAlaProTrpArgGluTrpGlu.....Ile 248
463 AAGGGCGCAATGACCTGATGAGTACGCAACAAACAGCGATTCCTCAT 512
249 GlnGlyArgGluAspAlaLleGluTyrAlaLysLysHisAsnValProVa 265
513 CCCGCTCACTCCCAAGAACCCGTGAGCATGATGAGACGTCATGACCA 562
265 lProValThrLysLysSerLleTyrSerArgAspArgAsnLeuTrpHisL 282
563 TCAGCTACAGGCTGGAATCTCGAGAACCCCAAGCAACCAAGCCCTTCA 612
.....

```



```

1 ORGANISM: Aspergillus fumigatus
US-09-675-784A-12974

Alignment_scores:
    Quality: 759.00      Length: 233
    Ratio: 4.059         Gaps: 0
    Percent Similarity: 80.258      Percent Identity: 60.944

Alignment block:
US-09-775-693-1 x US-09-675-784A-12974  ...

Align seg 1/1 to: US-09-675-784A-12974 from: 1 to: 257

7  AGCAAAAGCTCCGTGTTCTTCGCTTCCAGTGGCGGCGCTGGACACCTGTG 56
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
22 SerIysgIySvaIcYsIleuAlaPheSerIyGIyLeuAspIhrSerVa 38
57 CATCTCTCGTGGCTGTAAGAAAGGCTATAGCGTACGTCATGGCTATCGG 106
   ||||| ||||| ||||| .....||| ||||| .....||| ||||| .....|||
38 lIleuIystrPleuIleAspIugIyrgIuValValAlaIaPheIrrA 55
107 CCACATTTGGCCAGAGAAAGACTTGGAGAAACCAGAAAGAGGCACTG 156
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
55 IaaSnValGIyIglInclInuAspPheAspIleIargIuIySAlaIeu 71
157 AAGCTTGGGGCCAAAAGGTTCATTGAGAGATGTCAGGAGGAGATTGT 206
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
72 LysIleuGIyAlaValIyScYsGIuIleAlaAspLeuIArgIuIaPheI 88
207 GAGAGATTCATCTGGCGGCATCCAGTCCAGCGACATGTAGAGACC 256
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
88 eGIuInIleuCYsPheProAlaIleAlaCYsAsnAlaIleTYrgIuAsnV 105
257 GCTACCTCTCGGGACCTCTCTTCCGACGCCCTGCATCGCCGCCAAACA 306
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
105 aIyIreIleuGIyIhrSerIleuAlaIArgProValIleAlaIArgIaGI 121
307 GTGGAAATGCCCCAGCGGGAGGGGGCCAAAGTGTGTCACGCGCCAC 356
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
122 IlaSpIaIaIaIyStrGIuGIyCYsPheAlaIaIValSerIseIySrh 138
357 AGCAAAAGGGAAAGATCAGTCCGGTTTGAGCTCAGCTGCTACTACTG 406
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
138 rGIyIyGIyAsnAspGIuValArgPheGIuIleuAlaPheTYrAlaIeug 155
407 CCCCCCAGATAAGGTCATCTCTCCCTGGAGAGATCCCTGAATTCACAC 456
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
155 InProAspIleuYsValIleAlaIProIrrPaIyAspProAlaPheTYrAsn 171
457 CGGTTCAGGCGCGCATGACCTGTATGGAGTACGCCAAAGCAAGACAGGGAT 506
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
172 ArgPheAlaGIyArgSnsPleuIeuSerTYrIlaaIaGIuIySgIyI 188
507 TTCCATCCCGGTCACGCCCAAGACCCGTCGAGAGATGGATGAGAACTCA 556
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
188 eProValTYrSerThIySaIaIyStrPrrIerSerIeCYsPoiuAsnIeuA 205
557 TGCAATCAGCTACGAGGCTGGATCTTGGAGAAGCCCAAGAACCAAGCG 606
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
205 IahIscYsSerTYrGIuAlaGIyIleIeugIuAspProAsnValThPro 221
607 CCTCCAGGTCTTACACGAAGACCCAGGACCCAGCCAAAGCCCCACAC 656
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
222 ProIlaAspMetTYrPlySleuTYrGIuAspProIleuThrAlaIProAspGI 238
657 CCTCAGATTCTCGAGATCGAGTTCCAAAAGGGGTCCCTGTGAAGCTG 705
   |||||.....||| |||||.....||| |||||.....||| |||||.....|||
238 nProCIuAspPheThrAlaIhIspheGIuIhrGIyIleProValIySleu 254

seq_name: /cgn2_6/ptodata/2/paa/US09_NMW_COMB.pep:US-09-708-427-79359
seq_documentation_block:

```

```

; Sequence 79359, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708, 427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79359
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..308
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..308
; OTHER INFORMATION: Ceres Seq. ID 1965190
US-09-708-427-79359

```

```

alignment_scores:
  Quality: 655.50      Length: 305
  Ratio: 2.953        Gaps: 3
  Percent Similarity: 72.787   Percent Identity: 44.590

```

## alignment\_block:

```
US-09-775-693-1 x US-09-708-427-79359 ..
```

```
Align seg 1/1 to: US-09-708-427-79359 from: 1 to: 308
```

```

277 CTTGCCAGCCCTCATCGCCGCAACAAAGTGGAAATCGCCGAGGGGA 326
    :::::::::::::::::::::
1  MetAlaArgProValIleAlaIysAlaMetValAspValAlaIysGluVal 17
327 GGGGGCCCAAGTATGTGTCGCCAGCGCCACAGAGAAAGGGAGCATCAG 376
    |||||||
17  IGIAlaAspAlaValAlaHisGlyCysThrGlyIysGlyAsnAspGluVal 34
377 TCCGGTTGAGCTGAGCTGCTACCTACCTGCGCCGCAATTAAGTCATT 426
    |||||||
34  AlArgPheGluLeuThrPheThrAlaLeuAsnProGluLeuIysValVal 50
427 GCTCCCTGGAGAGTCCCTGAATTCCTAACCGGTTCAAGCCGCAATGA 476
    |||||||
51  AlaProThrParGluTrpAsp.....IleThrGlyArgGluAsn 63
477 CCGTAGGATGACGCAAGCAACACGCGATCCCATCCGCTCAGTCCCA 526
    |:::::::::::::::::::
63  PalAlaIleGluTrAlaIysIysHisAsnValProValProValSerLysL 80
527 AGAACCCGTGGACATGATGAGAACCTCATCATCATCAGTACGAGAGCT 576
    ||:::
80  ySerIleIySerIArgAspArgAsnLeuTrpHisLeuSerHisGluGly 96
577 GGAATCTTGAGAACCCCAAGAACAGCGCTCCAGGTCTCTACACGAA 626
    |||||||
97  AspIleLeuGluAspProAlaAsnGluProLysGluAspMetIyMeIe 113
627 GACCCGAGACCCGAGCAAGCCGCAACACCCCTGACATCTTCAGATCG 676
    ::::
113  tSerIleAlaProGluAsnAlaProSerIyProGluIyLeuGluIleG 130
677 AGTTCAAAAAAGGGTCCCTGAGAGTGACCAAGCTCAGATGGACACC 726
    ::::
130  IyIleIleAlaGlyAlProValSerIle...AsnGlyArgAsp..... 143
727 ACCCAACAGACCTCTCTGAGCTCTTCATGTACTGAACGAGTGGCGGG 776
    ::::
144  .....LeuSerProAlaSerLeuLeuAlaGluLeuAsnGluIleGlyG 158

```

```

777 CAAGCATGGCGTGGCCGCTATGTGACATGTGAGAACCCGCTTCATTGCA 826
    |||||||
158  IyHisIysGlyIleGlyAlaArgIleAspMetValGluAsnArgLeuValCysM 175
827 TGAAGTCCCGAGGTATCTACGAGACCCGCAAGCAGCACCCTTCATTAC 876
    |||||||
175  eIySerIArgGlyAlaIyGluTrIProGlyIyThrIleMetAlaAla 191
877 GCTCATTTAGACATGAGCGCTTCACCATGACCGGAGAGTGCGCAAAAT 926
    |||
192  AlaValArgIuLeuGluSerLeuThrIleuAspArgGluTrIleMetGlnTr 208
927 CAACAGGCGCTGGCTTGAATTTGCTGAGCTGTGATACCGGTTCAC 976
    ||:::
208  pIyAspIleIleAlaLeuIySyrAlaGluLeuValIyAlaGlyArgT 225
977 GGCCTAGCCCTGAGTGAATTTGTCGCCACACTGCATCGCCCAAGTCCAC 1026
    ::::
225  rPheAspProLeuArgGluInSerPheAspAlaPheMetGluIyValThr 241
1027 GAGCGAGTGAAGGCAAGTGCAGGTGCGCTCTCAAGGCGCAGGTGA 1076
    ::::
242  AlaThrThrThrGlySerValThrIleuIySerIySyrValAs 258
1077 CATCTCGCGCGGAGTCCCATCTCTCTCAATGAGAGACTGTGA 1126
    ::::
258  nValAlaSerArgIySerProIySerLeuIyArgGluAspIleSerS 275
1127 GCATGACGTCGACAGGTGATATGAGCCACATGACCGGCTTCATC 1176
    ||:::
275  erPheGluAsnGlyGluIleIyTrAsnGlnAlaAspAlaGluIyPheIle 291
1177 AACATCAATTCCTC 1191
    ::::
292  ArgLeuTrGlyLeu 296

```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW.COMB.pep:US-09-708-427-55744
```

```

seq_documentation_block:
; Sequence 55744, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708, 427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55744
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..309
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..309
; OTHER INFORMATION: Ceres Seq. ID 1935963
US-09-708-427-55744

```

```

alignment_scores:
  Quality: 639.50      Length: 301
  Ratio: 2.947        Gaps: 4
  Percent Similarity: 72.093   Percent Identity: 44.850

```

## alignment\_block:

```
US-09-775-693-1 x US-09-708-427-55744 ..
```

```
Align seg 1/1 to: US-09-708-427-55744 from: 1 to: 309
```

```

277 CTGGCCAGGCGCTGATCGCCCGCAACAGTGAATCGCCAGCGGGA 326
    ::::::::::::::::::::
1 MetalAargProvalIleAlaLysAlaMetValAspValAlaLysGluVal 17
327 GGGGGCCAAATATGTGTCCACGGCCCAAGAAAGGGAACATCAGG 376
    ::::::::::::::::::::
17 LcylAlaSpralaValAlaHisGlyCysThrGlyLysCysAsnSprGln 34
377 TCCGGTTGAGCTCAGCTGCTACAGCGGCCCCAGATTAAGTCAAT 426
    ::::::::::::::::::::
34 AlArgPheGluLeuThrPheTyrAlaLeuAsnProGluLeuLysVal 50
427 GCTCCCTGGAGGATGCTGAATTTCTACAAACGGTTCAAGGGCCGAATGA 476
    ::::::::::::::::::::
51 AlArgProTArgGluTArgPasp.....IleThrGlyArgGluAs 63
477 CCTGATGAGTACGCAAGCAACAGCGGATTCCTCCGGTCACTCCCA 526
    ::::::::::::::::::::
63 PalAlaLeuTyrAlaLysLysHisAsnValProValProValSerLysL 80
527 AGAACCCCGGAGCATGATGAGAACCTTCATGACATGACAGAGAGGT 576
    ::::::::::::::::::::
80 ySerLleTyrSerArgPaspArgAsnLeuTyrPHisLeuSerHisGluGly 96
577 GGAATCCTGAGAACCCCAAGCAACAGCGCTCCAGGTCTCTACACGAA 626
    ::::::::::::::::::::
97 AspLleLeuGluAspProAlaAsnGluProLysGluAspMetTyrMetMe 113
627 GACCCAGGACCAAGCAACAGCGCTCCAGGTCTCTACAGATCG 676
    ::::::::::::::::::::
113 tSerLleAlaProGluAsnAlaProSerLysProGluTyrLeuGluLleG 130
677 AGTTCAAAAAAGGGGTCCTGTAAGGTGACCAACAGTGAAGATGGACAC 726
    ::::::::::::::::::::
130 LylLleAlaGlyValProValSerLle...AsnGlyArgPasp..... 143
727 ACCCAGCAGACCTCTTGAGGCTTCATGACTGACGAAGAGTGGCGGG 776
    ::::::::::::::::::::
144 .....LeuSerProAlaSerLeuLeuAlaGluLeuAsnGluLleGlyL 158
777 CAAGCATGGCGTGGCGCTATGACATCGTGGAGAACCGCTTCATTTGAA 826
    ::::::::::::::::::::
158 LysHisGlyLleGlyArgLleAspMetValGluAsnArgLeuVal**M 175
827 TGAAGTCCGAGGTATCTACAGAGCCCGACAGGACCATCTTCATCAT 876
    ::::::::::::::::::::
175 eLysSerArgGlyValTyrGluThrProGlyGlyThrLleMetAlaAla 191
877 GCTCATTTAGACATCGAGGCTTCACCATGACCGGGAAGTGGCAAAAT 926
    ::::::::::::::::::::
192 AlaValArgGluLeuGluSerLeuThrLeuAspArgGluThrMetGlnTr 208
927 CAACACAGCGCTGGCTGAATTTCTGAGCTGGTGTATACCGGTTTAC 976
    ::::::::::::::::::::
208 pLysAspLleLleAlaLeuLysTyrAlaGluLeuValTyrAlaGlyArgT 225
977 GGCTGACCCCTGAGTGAATTTGTCGCGACATCGCCAGAGTCCCGAC 1026
    ::::::::::::::::::::
225 rPheAspProLeuArgGlnSerPheAspAlaPheMetGluLysValThr 241
1027 GACCGAGTGAAGCAAGTGAAGTGGTCCGTCCTCAAGGCGGAGTGTGA 1076
    ::::::::::::::::::::
242 AlaThrThrThrGlySerValThrLeuLysLysValTyrGlySerValAs 258
1077 CATTCCTGGCGGAGTCCCGACTGCTCTCATCAATGAGGAGAGTGTGA 1126
    ::::::::::::::::::::
258 nValAlaSerArgLysSerProTyrSerLeuTyrArgGluAspLleSers 275
1127 GCATGACAGTGCAGGGTATATGACCAACTGATGACACCGGTTTCATC 1176
    ::::::::::::::::::::
275 eRheGluAsnGlyGluLleTyr..AsnGln**ThrProArgLysLysSer 291

```

```

1177 A 1177
291 r 291
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-708-427-11259

```

```

seq_documentation_block:
; Sequence 11259, Application US/09708427

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: N. ALEXANDROV et al.

```

```

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

```

```

; FILE REFERENCE: 2750-1243P

```

```

; CURRENT APPLICATION NUMBER: US/09/708,427

```

```

; NUMBER OF SEQ ID NOS: 85364

```

```

; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 11259

```

```

; LENGTH: 283

```

```

; TYPE: PRT

```

```

; ORGANISM: Arabidopsis thaliana

```

```

; FEATURE:

```

```

; NAME/KEY: misc.feature

```

```

; LOCATION: 1..283

```

```

; NAME/KEY: misc.feature

```

```

; LOCATION: 1..283

```

```

; OTHER INFORMATION: Ceres Seq. ID 1822816

```

```

US-09-708-427-11259

```

```

align_scores:
Quality: 598.00 Length: 308
Ratio: 2.875 Gaps: 3
Percent Similarity: 67.532 Percent Identity: 40.260

```

```

alignment_block:
US-09-775-693-1 x US-09-708-427-11259 ..

```

```

Align seg 1/1 to: US-09-708-427-11259 from: 1 to: 283

```

```

277 CTGGCCAGGCGCTGATCGCCCGCAACAGTGAATCGCCAGCGGGA 326
    ::::::::::::::::::::
1 MetalAargProvalIleAlaLysAlaMetValAspValAlaLysGluVal 8
327 GGGGGCCAAATATGTGTCCACGGCCCAAGAAAGGGAACATCAGG 376
    ::::::::::::::::::::
9 .....V 9
377 TCCGGTTGAGCTCAGCTGCTACAGCGGCCCCAGATTAAGTCAAT 426
    ::::::::::::::::::::
9 AlArgPheGluLeuThrPhePheSerLeuAsnProGluLeuLysValVal 25
427 GCTCCCTGGAGGATGCTGAATTTCTACAAACGGTTCAAGGGCCGAATGA 476
    ::::::::::::::::::::
26 AlArgProTArgGluTArgPasp.....IleGlnGlyArgGluAs 38
477 CCTGATGAGTACGCAAGCAACAGCGGATTCCTCCGGTCACTCCCA 526
    ::::::::::::::::::::
38 PalAlaLeuTyrAlaLysLysHisAsnValProValProValSerLysL 55
527 AGAACCCCGGAGCATGATGAGAACCTTCATGACATGACAGAGGT 576
    ::::::::::::::::::::
55 ySerLleTyrSerArgPaspArgAsnLeuTyrPHisLeuSerHisGluGly 71
577 GGAATCCTGAGAACCCCAAGCAACAGCGCTCCAGGTCTCTACACGAA 626
    ::::::::::::::::::::
72 AspLleLeuGluAspProAlaAsnGluProLysGluAspMetTyrMetMe 88
627 GACCCAGGACCAAGCAACAGCGCTCCAGGTCTCTACAGATCG 676
    ::::::::::::::::::::
88 tSerValAspProGluAspProAlaProAspGlnProGluTyrLleGluLleG 105

```



Wed Feb 13 07:36:01 2002

us-09-775-693-1.rapn

Page 15

